Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

INGEGNERIA CIVILE, CHIMICA, AMBIENTALE E DEI MATERIALI

Ciclo 33

Settore Concorsuale: 08/B1 - GEOTECNICA

Settore Scientifico Disciplinare: ICAR/07 - GEOTECNICA

A METHODOLOGICAL APPROACH FOR THE PERFORMANCE OPTIMIZATION OF TRANSIENT SEEPAGE MODELS THROUGH INVERSE ANALYSIS

Presentata da: Ilaria Bertolini

Coordinatore Dottorato

Luca Vittuari

Supervisore

Guido Gottardi

Co-supervisori

Olivier Buzzi

Carmine Gerardo Gragnano

Esame finale anno 2021

Every challenging work needs self-efforts as well as guidance of elders especially those who are very close to our heart…

I dedicate this thesis to my family

Mum, Dad, Sara

for their constant support, encouragements and unconditional love.

I love you all dearly.

ABSTRACT

The topic of the Ph.D project focuses on the modelling of the soil-water dynamics inside an instrumented embankment section along Secchia River (Cavezzo (MO)) in the period from 2017 to 2018 and the quantification of the performance of the direct and indirect simulations in terms of observed-simulated behavior.
The commercial code Hydrus2D by Pc-Progress has been chosen to run the direct simulations. Different soil-hydraulic models have been adopted and compared.
The parameters of the different hydraulic models are calibrated using a local optimization method based on the Levenberg - Marquardt algorithm implemented in the Hydrus package. The indirect estimations of the model parameters are obtained from the minimization of an objective function starting from an initial set of the parameters and a dataset of observation points from in situ monitoring.
The calibration program is carried out using different types of dataset of observation points, different weighting distributions, different combinations of optimized parameters and different initial sets of parameters. The final goal is an in-depth study of the potentialities and limits of the inverse analysis when applied to a complex geotechnical problem as the case study (transient hydraulic and atmospheric boundary conditions, multilayered section, long period of simulation).
At first a qualitative then a quantitative methodology, by means of a set of proper indices/metrics, have been presented to investigate the behaviour of the large pool of indirect simulations in order to identify the optimized set/s of parameters and the hydraulic model that better simulates the bank behavior.
The predictive capabilities of the optimized models have been tested in the validation phase upon a new simulation period (2018-2019) and a quantitative estimation of the direct simulations performance is carried out similarly to what has been done in the calibration phase.

The second part of the research focuses on the effects of plant roots and soil-vegetation-atmosphere interaction on the spatial and temporal distribution of pore water pressure in soil. The investigated soil belongs to the West Charlestown Bypass embankment, part of the State Highway 23, Newcastle, Australia, that showed in the past years shallow instabilities and the use of long stem planting is intended to stabilize the slope. The chosen plant species is the Malaleuca Styphelioides, native of eastern Australia.
The research activity included the design and realization of a specific large scale apparatus for laboratory experiments. Local suction measurements at certain intervals of depth and radial distances from the root bulb are recorded within the vegetated soil mass under controlled boundary conditions. The experiments are then reproduced numerically using the commercial code Hydrus 2D that allows to simulate the plant root water uptake (RWU) contribution. Laboratory data are used to calibrate the RWU parameters and the parameters of the hydraulic model. In parallel, the physical and hydraulic properties of the investigated soil have been obtained by proper laboratory tests.

LIST OF CONTENTS

LIST OF FIGURES

GENERAL INTRODUCTION TO THE MAIN TOPICS OF THE RESEARCH (PARTS I AND II)

The research project focuses on two main topics:

- The modelling of the hydraulic and retention behaviour of an instrumented bank section along Secchia River (Cavezzo (MO)) under transient seepage conditions. The soil hydraulic properties for different hydraulic models are calibrated using an optimization algorithm based on the Levenberg-Marquardt method implemented in the commercial code Hydrus 2D by Pc-Progress. The calibration process is carried out considering various assumptions on observation datasets, different weighting distributions, combination of optimized parameters and initial guess of the parameters. The research is configured as an in-depth study of the potentialities and limits of the use of the inverse analysis method when applied to a complex geotechnical problem as the presented case study. A methodological approach for quantifying the quality of predictions of a large pool of indirect simulations is presented in order to identify the optimized set/s of parameters and the hydraulic model that better simulates the riverbank behaviour. The calibration and validation processes are firstly presented in theory then applied practically to the large set of available indirect simulations. The presented research project has been carried out in Bologna University, School of Engineering, under the supervision of prof. Guido Gottardi and dr Carmine Gerardo Gragnano.

The topic that has been briefly presented here will be discussed in every detail in PART I of the present thesis.

- The RMS project focuses on the investigation of the effects of the evapo-transpiration contribution of plants on the spatial and temporal distribution of pore water pressure in soil by means of the design and realization of a large-scale apparatus. The apparatus consists of a garden bed with a plant in the middle; sensors monitoring the soil water potential at certain depths and radial distance from the root bulb are installed. Different boundary conditions are imposed to the laboratory apparatus and experimental data are collected in time. The laboratory experiments are reproduced numerically using the commercial code Hydrus 2D by Pc-Progress. An in-depth calibration of the soil hydraulic properties by means of inverse analysis and the elaboration of a RWU spatial distribution initially by means of spatial distribution functions available in literature (Vrugt et al, 2001) then using experimental laboratory data are reported. Potentialities and limits in the use of available root spatial distribution functions are discussed. The presented research project has been carried out during the research period in the Priority Research Centre for Geotechnical Science and Engineering, University of Newcastle, Australia under the supervision of prof. Olivier Buzzi.

The topic that has been briefly presented here will be discussed in every detail in PART II of the present thesis.

Looking at similarities and differences between the two investigated case studies:

-both the topics are hydro-thermal-mechanical problems.

-the riverbank in PART I is subjected to unsaturated conditions as effect of river level fluctuations while the large scale apparatus in PART II is subjected to variable soil saturation subsequent to water level imposition.

-the riverbank is subjected to transient seepage conditions, the large scale apparatus when boundary conditions are applied is subjected to a transient seepage until the establishment of steady state conditions.

-for both the case studies, the mechanical and hydraulic parameters are dependent on soil saturation and for the large scale apparatus also on the root bulb influence.

-the riverbank section is a multi-layered domain and soil vertical and horizontal intrinsic heterogeneity is encountered within the same layer while the large scale apparatus is filled with homogeneous soil and a single layer has to be investigated. Riverbank soil heterogeneity is a topic not addresses in the research and layers are assumed as homogeneous considering the average of the hydraulic parameters obtained from laboratory tests for each layer.

-the installed sensors in the riverbank section collect monitoring data of water potential and water content in time, while in the large scale apparatus only water potential data are recorded.

-to model the behaviour of the riverbank section it is relevant to consider the soil hysteretic behaviour while for the RMS project the root water uptake by plant transpiration has to be computed with higher accuracy.

-in the large scale apparatus the boundary conditions are imposed (water level on the bottom boundary and relative humidity on the upper boundary). In the riverbank section boundary conditions are controlled (as the hydrometric water level of the river), partly controlled (atmospheric boundary conditions controlled by means of a meteorological station, simplifications are adopted for the computation of the evapotranspiration contribution), or uncontrolled (as in the bottom boundary of the model). From all this, we can understand the different complexity of the two problems: in the RMS project all the variable are controlled (temperature, humidity, suction, radiation, wind, water level imposed, lateral boundary conditions) due to the fact that is a laboratory experiment while in the riverbank project uncertainties on the boundary conditions are greater as in the majority of the in situ experiments.

-the evapo-transpiration contribution due to vegetation cover has been addressed in both the project but with a different degree of accuracy: in the RMS project the focus is on the plant activity so the degree of accuracy is extremely high while in the riverbank section the lack of experimental information leads to the adoption of simplifications.

-both the projects use input hydraulic parameters obtained by laboratory tests (evaporation tests) but obtained with different techniques: in the RMS project tensiometer and chilled mirror device (WP4 by Decagon Device) readings and the hand spry method have been used to investigate the SWRC while for the riverbank project the soil moisture release curve Hyprop by Meter and WP4.

-a calibration procedure has been applied to both the projects but with a different degree of complexity: the RMS project requires the optimization of one single layer (6 hydraulic parameters) while for the riverbank section the optimization of 4 layers (24 to 40 parameters). Given the lower complexity of the laboratory experiment of PART II and the lower number of indirect simulations performed, the evaluation

procedure used is simpler and quicker with respect to the one of PART I even if based on the same indices/metrics and the same base concepts.

-the calibration procedure applied to the riverbank project focuses on the optimization of the hydraulic parameters while in the RMS project on the optimization of hydraulic parameters and RWU spatial parameters.

-both the projects have as final goal the elaboration of a reliable pore water pressure distribution in certain time instants of the simulation that could be used as initial conditions for future stability analysis. Stability analysis could be performed on the investigated bank section along Secchia River and, for the RMS project, on the embankment along West Charlestown Bypass in Newcastle city which showed stability problems and long stem plantings have been proposed as in situ interventions. Stability analyses have not been performed in the present thesis and they stand as future developments of the research.

From this brief presentation, the strong connections between the two main topics of the research could be deduced. Both the topics are moving inside the same field, unsaturated soil mechanics and modelling of the water dynamics in transient conditions. While PART II focuses mainly on the simulation of the seepage phenomenon and the pwp distribution in close proximity of a root bulb to reproduce the conditions that usually occur in the first meter of soil, in PART I the focus is on the simulation of the pwp distribution in the riverbank body below the first meters of soil where the influence of the plants transpiration and evaporation from the soil surface are minimal. This is due to a lower number of information available from in situ monitoring that does not allow an accurate simulation of the evapo-transpirative contribution. Despite the fact that the chosen plant in PART II does not match the typical riparian vegetation in the Po Basin, the experimental setup and following procedure of RWU parameters calibration could be applied easily also to the case study of the riverbank section and its vegetative cover. The information obtained in the laboratory experiments could be integrated in a possible future numerical model of the investigated riverbank in terms of evapo-transpirative contribution in time to be applied uniformly on the vegetated atmospheric boundary or more precisely punctually in the domain in proximity of each riparian plant. The choice of which approach has to be used depends on the accuracy that is targeted. Finally, the two models, simulating the seepage in the bank body and the evapo-transpirative phenomenon close to the soil surface, together are able to capture in a very accurate way the pwp distribution in time of the investigated geotechnical problem (riverbank section subjected to transient boundary conditions). The pwp distribution, matching with accuracy the real monitored phenomena in the most interesting time instants of the simulation, could be imported as initial conditions for riverbank stability analyses which are the final purpose of these detailed numerical analyses. From all this, it is possible to understand that both the topics have the same ultimate goal and each of them has potentialities which balance the limits of the other.

PART I

1 INTRODUCTION TO PART I

River floods are a worldwide natural hazards with huge socio - economic impacts that are expected to rise in time due to population increase (urbanisation of flood-prone territory), economic growth (increase of property values), climate change (Tanoue et al, 2016), improvement in collecting and reporting flood disasters (Peduzzi et al, 2009) and less awareness of the population about natural risks. Flood, as defined by Directive 2007/60/EC (European Directive) is the temporary coverage of areas by water from rivers, streams, canals, lakes and sea for coastal areas. The present dissertation mainly focuses on floods from rivers and torrents.

The risk drivers used to understand and interpret these complex phenomena are ''*hazard*'', ''*exposure*'', ''*vulnerability*'' and ''*resilience*''. The summary definitions of these concepts are given in the following for sake of clarity.

The concept of ''*hazard*'' is linked to the possible occurrence of a phenomenon, process or human activity that may cause health impacts, injuries, loss of lives, environmental degradation and economic losses. Hazards could act alone or in combination, in one episode or more episodes sequential in time, and are defined by their location, origin, magnitude, frequency and probability. Hazard is often and incorrectly overlaid to the concept of risk, but it's currently recognized that hazard is only a factor of the overall risk.

''*Exposure*'' refers to people, property, infrastructure, housing, systems and all the human activities located in the hazard-prone areas that could be potentially affected. ''*Vulnerability*'' refers to the characteristic of a community, human activity, system or area that makes it susceptible of hazard damages. Vulnerability is multidimensional in its nature because it includes physical, social, economic, environmental, cultural and institutional factors. Often ''*exposure*'' is mistakenly associated with the term ''*vulnerability*''. Exposure is a necessary but not sufficient factor of the risk. It is possible to be exposed but not vulnerable (e.g. adopting mitigation measures that zero the vulnerability in exposed areas susceptible of hazards).

The ''*resilience*'' is the ability of a community, society, system to recover from the consequences of an hazard in a short time and in an efficient manner restarting vital infrastructures and services. All the controlling risk factors (''*hazard*'', ''*exposure*'', ''*vulnerability*'') change temporally and geographically and a huge effort has been done in the past years to assess the spatial-temporal characteristic of the flood hazard under actual and possible future climatic scenarios at the global scale (Tanoue et al, 2016, Ward et al, 2013, Hirabayashiet al, 2013, Winsemius et al,2016, among others). The final goal is the elaboration of a global risk map to guide disaster risk reduction policies aimed at preventing new disaster risks and reducing existing ones, contributing to resilience strengthening and a future sustainable development of countries.

Italy has experienced a strong delay in the promulgation of norms addressed to natural hazard risk reduction (landslides and floods) in urban planning. Law n° 183 (May,18,1989) is the first national regulation that identifies the hydrographic basin as a territorial entity to be addressed and coordinated as a whole by the Basin Authority by means of a Basin Plan. Despite that, the law has full implementation only after the catastrophic event of Sarno (May,5,1998). L. 267/1998 gives acceleration on the individuation and classification of the areas subjected to hydrogeological hazards and the elaboration of PAI (Plans for the

Hydrogeological Asset of Basins). Directive 2007/60/CE (Floods directive - FD) implemented in Italy with the Decree 49 (23/2/2010) defined new common lows across Europe for the evaluation and management of the flood risk introducing three hydraulic hazard and risk scenarios and the Management Plans of the flood risk.

In 2008, the Italian Institute for Environmental Protection and Research (ISPRA) has been established (Decree n° 112 of 25 June 2008). In 2015, in the field of the hydrogeological risk, ISPRA mapped the entire national territory using the information collected by Basin authorities, Regions and independent provinces according to three scenarios (P1,P2,P3).

In 2017 ISPRA updated the flood hazard map of the national territory using the same different risk scenarios (Trigila et al, 2018): P3, with return time between 20 and 50 years (frequent floods), P2 with a return time between 100 and 200 years (average probability of flood occurrence) and P1 (low probability of floods or extreme event scenarios). The conclusions that could be drawn from this study are extremely alarming: 4,1% of the national territory has a P3 hydraulic hazard, 8,4% a P2 and 10,9% a P1 flood probability. Emilia Romagna, region in the northern Italy, has a territorial extension of 22452 km^2 and possesses the Italian primacy of the highest rate of territory in danger (11,1%P3;45,7%P2, 35,5%P1) (Trigila et al, 2018; Trigila et al, 2015). The considerable extension of area with P2 hazard in E.R. is connected to the main and secondary natural hydrographic network and the dense network of artificial drainage channels (Trigila et al, 2015).

To meet the need to better manage the hydrogeological risk, Italian public bodies were established in order to be in charge of large river basins, carrying out infrastructure maintenance and risk mitigation interventions, funding research activities for early warning systems and emergency plans. AIPo (the interregional agency for the Po River) deals with the management of the Po River and its basin which covers a huge surface (74 000 km^2) divided between six different Italian regions and host a population of about 16 million of people.

1_1 AN OVERVIEW OF THE PROBLEM

The construction of embankments along river banks or costal area is a very old practice probably dating back to the dawn of civilization. The main function of these earthen retaining structures is to limit a seepage process through it, which is guaranteed by a low or very low permeability of the construction materials used. For this reason fine-grained materials with a low or very low fraction of coarse-grain material are used. In the majority of the cases, for the construction of artificial embankments, the material chosen is the one available on site, in order to limit the cost of the intervention. To increase the levee performance, sometimes a cover or an inner layer with better characteristic is chosen and a compaction phase is used to reduce the hydraulic permeability of the soil. A natural slope has to be addressed with even more care due to the higher spatial variability that is likely to be encountered in the inner layers. Even within a so-called '' homogeneous layer'', soil properties usually show certain variability.

Compacted soil, typical of river embankments and transportation infrastructures, is unsaturated in the construction phase, but due to microstructure changes (loading), hydrometric fluctuation of the river level and weather and vegetation conditions (rainfall, evaporation and transpiration contributions) their degree of saturation could change considerably in time (Bicalho et al, 2018). The same conclusion could be extended to natural soils. Since the key parameter in slope stability is the soil strength which varies with

suction, water content and degree of saturation, it is of primary importance to have a good and reliable estimation of the spatial and temporal distribution of these variables (Rocchi et al, 2018a). However this problem seems yet a huge task for research and standard geotechnical studies because it requires site monitoring and advanced laboratory tests (Gragnano et al, 2018). The development of high capacity tensiometers (Mendes et al, 2008) and the use of instrumentations developed for agricultural purposes for the monitoring of superficial soils (Bittelli, 2011) and applied to monitor deeper soil layers are useful tools to achieve this goal (Rocchi et al, 2018b).

As suggested by Gottardi et al (2016), initial and boundary conditions have to be assessed properly because they can influence strongly global safety conditions. The design of earthen structures is entrusted to stability analyses based on simplified hypothesis such as steady state conditions associated with the expected flooding peak and the limiting situations of dry and total saturation during rapid drawdown of the river (Dapporto et al, 2003; Calabresi et al, 2013), disregarding issues related to soil partial saturation. All these assumptions lead to over-conservative results in terms of probability of failure (Gottardi and Gragnano, 2016). The assumption of steady state conditions equal to the maximum peak expected could be reasonable for coarse-grained embankment but not for fine-grained soil for which the penetration of the saturation line is poor and the embankment remains mostly in unsaturated conditions (Calabresi et al, 2013). For this reason, fine-grained soil embankment fears persistent flooding event of moderate water level instead of high water level occurred in a short period of time (Calabresi et al, 2013). Moreover, Rinaldi et al (2004) underlines the importance of taking into consideration in bank stability analysis the complex interaction between confining pressure given by the river and the pore water distribution in the bank. Another crucial aspect is the estimation of the transpiration contribution given by plants and the root reinforcement effect on slope stability due to an increase in soil shear strength.

The global picture that comes out from the present literary review is of a very complex geotechnical problem, with a number of variable aspects to take into consideration for a reliable stability analysis elaboration. Among other aspects the highly complex relation between infiltration, water content and pore pressure change that is non-linear and hysteretic (Toll et al, 2016).

Realistic hypothesis on initial conditions and on transient boundary conditions (transient seepage conditions) together with an in depth study of unsaturated and heterogeneous soils should be at the base of any reliable risk susceptibility analysis (Gottardi and Gragnano, 2016).

Recently, a consistent number of studies has focused on positive and negative pore water pressure (Simon et al, 2000; Rinaldi and Casagli, 1999 among others) incorporating their contributions in banks stability. These new discoveries allow reconsidering the simplified hypothesis previously adopted in various bank stability methods (Rinaldi et al, 2004). In practice, slope stability is assessed using numerical analysis performed by means of Limit Equilibrium Analysis, Finite or Discrete Element Methods. Finite Element Methods are used to determine the solution in terms of pore water pressure and water content distribution calculated for each node of a specific domain.

1_2 PROJECT BACKGROUND AND AIMS

The evaluation of model behaviour and performance is performed commonly comparing observed and simulated data and investigating the closeness between them. Observed data are obtained by field measurements made within the domain.

But, the general approach, based on the inspection simulated-observed data, allows the modeller to formulate a subjective assessment of the model behaviour, usually dependent on the systematic or timing over or under-prediction of the model. To evaluate objectively model behaviour, it is necessary to make a mathematical estimate of the error between simulated and observed data. The majority of the evaluations of model behaviour performed in literature are carried out presenting only graphical comparison and subjective assessments. The diagnosis that can be made from a graph need to be supported by quantitative measures.

Statistical metrics such as efficiency criteria, accuracy criteria, model selection criteria, statistical significance criteria represent powerful means to perform vigorously a model evaluation. The importance of using a large number of metrics has been already stated in different research field but there is a general lack of guidance in how to select the actual metrics to use. Each metric has specific pros and cons which have to be taken into consideration during model evaluation. In fact different metrics put emphasis on different systematic or dynamic behavioural error of the model. It is therefore important to compare the relative importance and suitability of each of these techniques to evaluate model prediction.

Several reasons lead to evaluate the performance of the model using sets of metrics/indices:

(1) It is necessary to estimate quantitatively the model ability to reproduce historic data. In other words, it is necessary to evaluate the performance of a **direct simulation** based on observation points (laboratory or in situ data). The use of a set of metrics/indices in these cases gives insights on:

- The type of data that are better or worse represented by the model (for example suction or water content information).
- In which temporal periods of the simulation the phenomena are better or worse represented (for example which flood events, which off-peak periods..) in order to focus the effort of a calibration phase on those periods of greater interest for the purposes of a certain study.
- Which observation points (sensors) are better or worse represented by the model in order to understand if the problem is related to the reliability and accuracy of the recorded data or on the ability of the model to reproduce them.
- Which parts of the domain (for example layers) are better or worse represented by the model in order to understand where to concentrate the efforts of the calibration phase.

(2) It is necessary to compare the same **direct simulation** under different constitutive models representing the soil behaviour (for example different hydraulic models or different mechanical constitutive models), in order to select the one/s that is/are able to represent optimally historic data.

(3) It is necessary to evaluate the performance of **indirect simulations** (inverse analysis) in order to select the optimized set/s of parameters that enhance the performance of the direct simulation from which the calibration phase has started.

(4) It is necessary to evaluate improvements in the **modelling approach** obtained performing structural modification to the model (different from changes in the model parameters considered in point (2)), including newly and additional observation information and representing important spatial and temporal characteristics.

(5) It is necessary to estimate quantitatively the model ability to reproduce future data (the predictive capability of the model). In other words, it is necessary to evaluate the performance of a **direct simulation** during the validation phase.

(6) It is necessary to compare the actual model effort/performance with previously done studies.

The aims of the present study are the following:

- To perform a quantitative evaluation of a **direct simulation** of the investigated phenomena (flooding history of the year 2018) using different hydraulic models, using different typologies of recorded data (pressure head or water content), different observation points (sensors) and trying to concentrate the attention on the behaviour of the simulation in the different temporal instants and locations of the modelling domain. Points (1) and (2) of the previous list.
- To elaborate and evaluate a protocol for analyzing the performance of **indirect simulations** and for guiding the **calibration phase** of the investigated simulation by means of a quantitative and objective procedure.
- To perform a quantitative evaluation of a **direct simulation** on a newly investigated period (flooding history of the year 2019) using the different set/s of optimized hydraulic parameters selected in the previous calibration phase. In fact, a good calibration procedure is necessary to have a robust and flexible model whose predictive capacities have to be tested in a proper **validation phase**. Point (5) of the previous list. As stated by Thirel et al (2015), calibration works to accommodate reality but is important that it doesn't remain ''*a mathematical marionette able to dance to a tune it has already heard*''.

The present work stands as an interesting case study of application of **inverse analysis** using the commercial code Hydrus 2D by PC-Progress. Hydrus 2D is a powerful numerical model that can be used for both direct problems when the initial and boundary conditions for all involved processes and corresponding model parameters are known, as well as inverse problems when some of the parameters need to be calibrated or estimated from observed data.

All these aspects are dealt in a problem of high complexity (multi-layered domain, transient phenomena). The topics are addressed by means of an evaluation and critical analysis of the indirect simulations output using different sets of statistical metrics/indices and different types of plots/tables which are able to highlight different aspects of the performed indirect simulations. The author believes that this particular case study is able to help future modellers to analyse quantitatively their model outputs through a clear framework because pros and cons, difficulties and potentialities encountered in a parameters optimization procedure using the commercial code Hydrus 2D when dealing with a complex problem, are encountered and faced.

The present work is subdivided in the following chapters:

In chapter §1 and §2 the main theoretical background on which the research work is based is presented and examined in detail. In particular chapter §1 highlights the relevance of the topic ''protection of the territory'' together with its main social and economic impacts and how this problem has been addressed by the Italian and European governments. Moreover the meaning of the main risk drivers (*hazard*, *exposure*, *vulnerability* and *resilience*) has been presented and discussed. In chapter §1_1, the topic of the embankment stability has been exposed highlighting the difficulties and complexity of the problem to be described which includes the description of many phenomena such as the transpiration contribution of plants, the evaporation contribution from the soil surface, rainfall infiltration and infiltration and seepage due to river hydraulic level variations and animal activity (burrows formed in the river body). Complexity could be found also in the model parameters that change in space and time due to soil heterogeneity and soil hysteretic behaviour. The first attempts made in the past to address the problem using simplified hypothesis together with the new attempts to consider the phenomenon in its whole complexity are discussed.

Chapter §2 contains the literary review with the definition of the main soil properties as the matric suction and the osmotic suction (chapter §2_1), the definition of the SWRC and the hydraulic conductivity function (chapter §2_2) and the main hydraulic models to describe the SWRC as the van Genuchten model (chapter §2_3); the VGM considering an air-entry value of 2cm (chapter §2_4), the MVG of Vogel & Cislerova (1988) (subchapter §2_4_1) and the MVG of Vogel et al (2000) (chapter §2_4_2) and the VGM considering the hysteretic behaviour (both the parameters of the main drying and the main wetting curve enter in the SWRC equation). Chapter §2_7 presents the Richards equation in its different forms and the intrinsic problems related to its resolution are discussed. In chapter §2_7 the commercial code Hydrus 2D is presented with its main potentialities in the direct and indirect (inverse analysis) problem. In sub-chapter §2_7_2, the full set of criteria/indices that are used in the following are presented giving details on pros and cons of their use for model performance evaluation divided in the main typologies: accuracy measures (§2_7_2_1); efficiency criteria (§2_7_2_2); Model selection criteria (§2_7_2_3) and statistical significance criteria (§2_7_2_4).

Chapter §3 presents the case study together with the individuation of the geographic location of the instrumented bank section, the geometry of the section and the subdivision in layers performed through CPTUs interpretation and the results of the physical characterization of the different layers. In chapter §3_1, the installed sensors are presented in terms of position in the bank section, accuracy and calibration procedure adopted. Information are given also on the installation procedure adopted and the main problems encountered. Chapter §3_2 focuses on the laboratory procedure adopted to characterize hydraulically the investigated soils and the hydraulic parameters obtained from laboratory SWRC are presented together with a statistical analysis of the results.

Chapter §4 discusses the preliminary tests performed on the different metrics/indices presented in chapter §2_7_2 in order to individuate which show a logic and reliable behaviour in the different hypothesized cases and which show ambiguous behaviour. Metrics are tested using synthetic data (of water content and pressure head) with different degree of noise. Possible explanations of the different behaviour of the metrics/indices are given.

Chapter §5 presents the numerical model of the investigated bank section elaborated for the simulation of the 2018 period. Details on initial conditions, boundary conditions, chosen observation points, model parameters, time discretization, spatial discretization, iteration criteria are given. Subchapter §5_1 presents

the sensitivity analysis performed on the hydraulic parameters of the models (the hysteretic ones and the non-hysteretic ones), that is able to give precious information to be used in the following calibration phases.

In chapter §6 a statistical comparison of the performance of models simulating the 2018 year using different hydraulic models (VGM, HVGM, VGM-2cm, HVGM-2cm) by means at the beginning of qualitative graphical methods then quantitative methods (set of metrics/indices), has been presented. A set of statistical indices has been applied to the whole set of observation points to give a rough idea of the performance of the simulations during the whole investigated period, then it has been applied to interesting time periods (major flood peaks), then to single observation points in order to isolate the behaviour of different zone of the domain. Together these information are able to guide the modeller in the difficult process of optimization of the model performance, identifying where the criticalities are or helping in the choice of the best model to represent a certain soil behaviour. Chapter §6_2 focuses on the performance of the VGM, the simpler hydraulic model among the considered ones. Analysis for single observation points/single sensor, whole set of observation points /sensors, different typologies of observation points (pressure head data, water content data or both), different time intervals (peak or off-peak periods) are presented in detail.

In chapter §7 and its subchapters, a calibration protocol divided in three phases will be firstly explained then applied to the investigated models in order to select from a large pool of indirect simulations the final datasets of potentially optimized parameters. Chapter §7 focuses on the presentation of the calibration programme performed on the four direct models (HVGM, VGM, VGM-2cm and HVGM-2cm), explaining the different datasets of observation points used, the different distributions of weights given to the observation datasets, the different combinations of hydraulic parameters to be optimized which could belong to the same layer or group of layers. Subchapter §7_1 presents a preliminary analysis of the results of the calibration programme performed. The percentage of successful simulations is investigated and put in relation with the hydraulic model used, the observation dataset used, the number of hydraulic parameters etc..

In subchapter §7_2 a preliminary quantification of the performance of the groups of inverse analysis (which are composed of indirect simulations optimizing the same layer/s using the same datasets of observation points and the same weighting distribution) by means of a proper set of metrics/indices is presented; the quantification is operated on the whole 2018 simulation period. This preliminary analysis is used to exclude the groups of inverse analysis that behave poorly in order to reduce the high number of simulations that have to be analysed one by one in the next phase of the calibration (first phase of the calibration programme). In this sense, subchapter §7_2 could be considered a preliminary and rapid phase of evaluation of the performance of groups of inverse analysis.

In subchapter §7_3 the first phase of the calibration programme is presented and then applied to the present case study. The first phase uses a set of metrics/indices to evaluate the deviation between observed-simulated datasets. The set of metrics have been applied for the evaluation of the performance during the whole simulation period and during the flood peak and off-peak periods. The chosen metrics are of typologies accuracy, efficiency, model selection criteria and statistical significance criteria. Points are given to the values of the metrics, calculated for each indirect simulation, based on the distance from the best fit of the index. Points of each metric are scaled based on an internal weight with respect to the other metric of the same group (accuracy, efficiency etc) and based on an external weight with respect to the

other groups of metrics. When the performance is calculated for each peak periods, points are scaled based on the persistence of the flood and the hydrometric level reached.

In subchapter §7_4, the second phase of the calibration programme is presented. The focus is switched on the results of the inverse analysis, the obtained optimized parameters and on the last iteration of the inverse analysis.

In subchapter §7_5, the third phase of the calibration programme is applied to the sets of indirect simulations that have overcome the first and the second phases. The attention is focused on the performance of the single sensor considering the whole simulation period and the two main flood events of the 2018, while in the first phase of the calibration the performance of the indirect simulations was investigated based on the performance of the whole set of available observation points on the whole simulation period and on the flood peak/off peaks periods. The set of chosen metrics is composed of accuracy and efficiency metrics. Weights are given to the values of the metrics according to the different positions in the bank section, the different type of measurement (PH or wc) and points are calculated based on the enhancement with respect to the base simulation (which uses the set of initial parameters).

In subchapter §7_6 inverse analysis to optimize the hydraulic parameters of the embankment layer are carried out starting from a different set of initial parameters (with respect to the one used in the previous calibration programme performed) in order to investigate its role in the indirect problem and the behaviour of the objective function starting from a different point in the parameters space.

In subchapter §7_7 weighting distributions (different from the ones used in the previous calibration programme) are applied to the dataset of observation points used in the inverse analysis to optimize the hydraulic parameters of the embankment layer. Weighting distributions are based on the different positions of the sensors in the bank section, the different type of sensors and the different temporal instants in which the observed data are recorded.

In chapter §8 the procedure of performance evaluation in three phases used for the calibration programme has been applied to a new monitoring period (year 2019) in order to evaluate the forecasting capability of the investigated models (validation phase). The sets of optimized parameters that have overcome the third phase of the calibration are tested again in order to individuate the final set of optimized parameter to use for the investigated bank section.

In chapter §9 the add-on module SLOPECUBE by Pc-Progress has been used to elaborate the Scalar Field of the Local Factor of Safety (SFLFS) of the investigated bank section (2018 year). Differences in the Local Factor of Safety between the model using the set of base parameters and the model using the set of optimized parameters are computed for different vertical sections and for different time instants of the simulation. Stability analysis are not the focus of the present work that locates himself on a previous step, the elaboration of reliable pwp distributions to be imported as initial conditions for future stability analysis. For this reason, as a conclusion of the work, it is considered interesting to highlight the important and immediate repercussions on the results of simplified stability considerations adopting an optimized set of parameters (obtained after an in-depth calibration and validation process) instead of the initial set of hydraulic parameters.

In chapter §10 final remarks on the performed work together with possible future developments are presented.

2 LITERATURE REVIEW

The literary review of the present thesis has to be intended as the background of information to all the subsequent work. An overview on unsaturated soils, their hydraulic and mechanical properties and the water retention models of interest is presented.

The understanding of unsaturated soil mechanics principles has a wide range of interests because it concerns all geotechnical problems associated with soil above water level and compacted soils. The classical soil mechanics assumes that soil voids are filled only by liquid phase (saturated conditions) or only by gaseous phase (dry conditions) and their mechanical behavior could be described by Terzaghi low for effective stress. These two limit conditions (saturated-dry) are addressed as singular cases in a wide range of saturation ratio that soil could assume. In unsaturated conditions the phases are four: air, water, solid phases and the contractile skin or air-water interface (Fredlund et al, 1978) that pulls soil particles together through surface tension like an elastic membrane. The pore-air pressure (u_a) is taken as reference for the independent stress variables, net normal stress (sigma- u_a) and matric suction (u_a-u_w) where u_w is the pore-water pressure. For geotechnical engineering purposes pore air pressure could be considered equal to the atmospheric pressure and set equal to zero (u_a=0).

2_1 TOTAL SUCTION, MATRIC SUCTION AND OSMOTIC SUCTION

It is relevant at this stage to give a formal definition of the property ''suction'' as the negative of soil water potential Ψ_{tot} (per unit volume). The soil water potential is the driving force that governs water movement in soil and it is formally defined as ''*the amount of work that must be done per unit quantity of pure water in order to transport reversibly and isothermally to the soil water at a considered point, an infinitesimal quantity of water from a reference pool*'' (Aslyng, 1963). The reference pool of water is arbitrary chosen at zero water potential. Soil is subjected to a variety of possible forces that influence water movement that are the result of interactions between soil matrix and water (e.g. gravity, presence of solutes, external gas pressure etc). Briefly, suction could be seen as the result of two components [equation 1]:

$$suction = -\Psi_{tot} = \Psi_o + \Psi_m \qquad \text{[equation 1]}$$

where Ψ_o is the osmotic suction and Ψ_m is the matric suction (both in units of pressure).

Osmotic suction is due to solutes dispersed in bulk water. As the concentration of ions increases, the osmotic suction increases. This component of the equation could play a central role in problems with physical barriers (membranes) to solute movement or water movement via vapor. Matric suction is generated by two phenomena: the capillarity phenomenon associated with the surface tension at the interface between air and water in soil pores and the adsorption (Lu and Likos, 2004). Capillary suction is based on the same principle of capillary rise in a tube. Water is a dipole molecule that inside a water phase is perfectly balanced (net forces equal to 0). At the interface between water and air, the system is not equilibrated (net force toward the water surface). To restore the equilibrium water molecules are arranged in such a way to create a tensile pull called surface tension. The surface tension acts like an elastic membrane in tension. When in contact with a solid surface as the borders of the cylinder, the interface assumes a concave curvature (Tarantino, 2013). The angle formed between water and surface is lower than 90° when adhesive forces (between water and soil) prevail on cohesive forces in the liquid phase. If menisci are present between soil particles, an additional compressive force is experienced by particles. The result is

an overall increase in soil shear strength. The absorption component of suction is the contribution to the water potential given by the attraction experienced by water molecules to opposite charged soil surfaces or ions attached to surfaces. The principles governing absorption component result from electrical field. The strength of the bounds between particles and surface are inversely proportional to the distance between them.

Changes in soil water content correspond to changes in matric suction. To a lower value of water content corresponds a higher value of matric suction and less water-filled spaces. Another important consequence is that soil hydraulic conductivity decreases with water content since water pathways are more tortuous, narrower and fewer and the large pores are the ones that drain first (Youngs, 2005). The change of permeability with matric suction is called ''permeability function'' and it expresses the decrease of conductive area of water to the water flow due to the increasing quantity of air trapped in the voids. The movement of water along thin films of absorbed water around soil particles is the primary mechanism of flow (Tokunaga, 2009).

2_2 THE SOIL WATER RETENTION CURVE AND HYDRAULIC CONDUCTIVITY FUNCTION

The relation between suction and volumetric water content (or gravimetric water content or saturation ratio) is described by soil water retention curve (SWRC) which exhibits hysteresis with respect to the wetting/drying history. Note that many other different names are given to the same relation (suction-VWC) as SWCC (Soil Water Characteristic Curve) and WRC (Water Retention Function). The mechanisms that assure the hysteretic response are broadly recognized as the ink-bottle effect due to non-uniformity in interconnected pores, changes in pore structure due to wetting and drying, air entrapment, capillary condensation and aging effect dependent on drying and wetting paths (Hillel, 1980; Lu and Likos, 2004; Likos et al, 2014). A similar hysteretic behavior is observed in the Hydraulic Conductivity Function (HCF) that relates hydraulic conductivity to suction. To simplify seepage analysis, hysteretic behavior is commonly disregarded, leading to inaccurate predictions of pore water pressure and volumetric water content distribution (Liu et al, 2016).

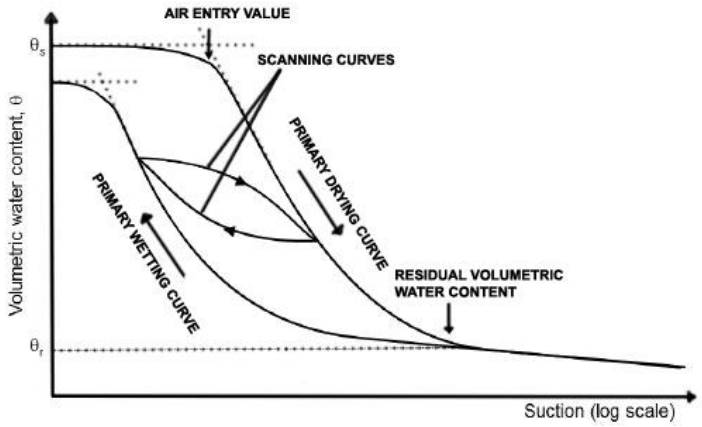


Figure 1 SWRC after Toll, 2012. The main wetting curve and the main drying curve are represented together with the graphical indication of the parameter α (air entry value) and θr (residual water content).

In Figure 1 a typical response to wetting and drying cycles of an initially saturated sample is presented. Upon drying the material maintains a fully saturated state until a threshold value called air-entry value

(AEV) is reached. It is generally considered as the graphical point corresponding to the intersection of tangents. The finer the material, the higher the air-entry value (AEV). AEV could be seen as the matric suction able to break menisci formed by the water surface tension in the large pores, the ones drained first (Rahardjo, 2019). Exceeded the air-entry value, air starts to penetrate in the voids (quasi-saturated soil state). As drying proceeds, the saturation ratio decreases while suction increases, following the so-called ''main drying curve''. In a partially saturated soil state, the desaturation is rapid and obtained with a relatively low change in matrix suction until a residual degree of saturation (or a residual water content) is reached. At this point it is necessary to increase significantly the suction value in order to change the saturation ratio (residual state). The residual water content is considered to occur at the intersection of tangents. Theoretically the maximum matric suction value is not fixed, however a value of 1000 Mpa, that corresponds to a water content of 0,06%, has been adopted for thermodynamic reason (Fredlund et al, 2012). If at that point the specimen is wetted, the degree of saturation increases and another path is followed known as ''main wetting curve''. The main drying curve and the main wetting curve do not coincide. If the wetting is extremely slow a full saturation could be reached, while if it is rapid, air could be trapped in the pores and a degree of saturation equal to 100% is not obtainable. If during the wetting/drying path the process is reversed another path is followed internally to the zone between the two main curves. This path is called ''scanning curve''. All the intermediate values of water content within the zone individuated by the two main curves could be assumed due to the hysteretic behavior already mentioned. SWRC is dependent on various factors such as pore size distribution, shape, tortuosity, specific surface area, mineralogy of soil particles.

2_3 THE VG MODEL

SWRC are continuous and smooth curves obtained fitting a finite number of laboratory points using numerical models. The van Genuchten (1980) analytical model gives a good approximation of experimental results for many soil types. VG equations [see equations 2,3,4] are a revision and modification of the power-law function proposed by Brooks and Corey (1964) that introduced an air-entry value above which the soil could be considered saturated.

$$S = \frac{\theta - \theta_R}{\theta_S - \theta_R} = \frac{1}{[1+(\alpha|h|^n)]^m} \quad for\ h < 0 \qquad \text{[equation 2]}$$

$$S = 1\ for\ h \geq 0 \qquad \text{[equation 3]}$$

$$h = \frac{s}{\gamma_w} = \frac{(u_a - u_w)}{\gamma_w} \qquad \text{[equation 4]}$$

Where S is the effective saturation degree, θ, θs, θr are the current, maximum and residual water content respectively, α is the inverse of the air-entry value (kPa^{-1}), h is the suction head, s is the matrix suction. The parameter α represents the inflection point, n defines the slope of the water retention curve. “m” is a fitting parameter identical for the main wetting and drying curves, that could be approximated by the expression in equation 5 (Mualem, 1976):

$$m = 1 - 1/n \qquad \text{[equation 5]}$$

This constraint to the parameter “n” is widely used in literature in order to reduce the number of unknown parameters of the model. Moreover if “m” and “n” are independent parameters, the integral has no

analytical solution and could be solved only adding mathematical complexity (Bittelli et al, 2015), while assuming “m” equal to “n” permits a direct integration of the Mualem’s predictive model.

The parameter ‘’n’’ is related to the pore-size distribution of the soil while the parameter “m” is related to the overall symmetry of the SWRC (Likos et al, 2014). Soils with a narrow pore-size distribution (and for this reason a narrow suction range) have relatively large values of ‘’n’’ compared to soil with wider pore-size distribution. The hydraulic conductivity of unsaturated soils is function of the degree of saturation S and it could be obtained from VG expression combining it with the statistical pore-size distribution of Mualem (1976) [see equation 6]. Reliable and accurate measurements of hydraulic conductivity are time consuming, costly and complex, all this leads to the necessity of developing indirect methods to obtain K(h) (Vogel et al, 2000).

$$K = K_s\sqrt{S}\,[1-(1-S^{\left(\frac{n}{n-1}\right)})^m]^2 \qquad \text{[equation 6]}$$

Where K_s is the saturated hydraulic conductivity.

2_4 THE VG MODEL IMPOSING A -2cm AIR ENTRY VALUE

The imposition of an air entry value of -2 cm to the van Genuchten’s formulation is recommended for fine-textured soils (clay and silty clays) which have an ‘’n’’ parameter close to its lower limit (1). This correction introduces a small air entry value that has no influence on SWRC itself but only on the hydraulic conductivity function close to saturation. An air entry value of -2cm forces the slope of the retention curve close to saturation to be zero (dθ/dh=0) and, as a result, the K(h) function is less non-linear and more accurate close to saturation. This correction is absolutely not recommended for medium and coarse textured soils for which the α parameter become relatively large (>0,02-0,05 cm). Vogel and Cislerova (1988) and later Vogel et al (2000) were the first to underline the problem of the K(h) function close to saturation.

2_4_1 MVG ACCORDING TO VOGEL AND CISLEROVA (1988)

Ippish et al (2006) demonstrated that under certain conditions ($n<2$ and $\alpha h_a>1$ where h_a is the air-entry value) the van Genuchten (VG) model predict erroneous hydraulic conductivities (Bittelli et al, 2015). Some fine-grained materials interpreted with VG model show non-linearity close to saturation and this aspect could cause instability, lower accuracy and problems of converge in the FE models. The permeability function is extremely sensible to change in the slope of SWRC, small changes of the function SWRC close to saturation could have big impacts on convergence of the numerical simulation, especially when fine-textured soils are investigated (Vogel et al, 2000).

Vogel and Cislerova (1988) proposed a modified set of equations of the VG model (Modified van Genuchten model MVG) in order to overcome the possible uncertainties in the determination of h(θ) near saturation. The physical parameter θ_s is replaced by the parameter θ_m with no physical meaning ($\theta_m \geq \theta_s$). The resulting SWRC is expressed with two distinct equations: from h=0 to $h=h(\theta_s)$ a linear equation, for $h>h(\theta_s)$ a non-linear equation. The hydraulic conductivity function is obtained substituting into Mualem’s relation the modified expression of the SWRC and then integrating (Vogel and Cislerova, 1988). The Modified van Genuchten model proposed by Vogel and Cislerova (1988) could be written as follows [equation from 7 to 11]:

$$\theta(h) = \begin{cases} \theta r + \frac{\theta_m - \theta_r}{[1+(-\alpha h^n)]^m} & h < h_s \\ \theta s \quad h \geq h_s \end{cases}$$ [equation 7]

$$K(h) = \begin{cases} K_s \; K_r(h) & h < h_s \\ Ks & h \geq h_s \end{cases}$$ [equation 8]

where

$$K_r(S_e) = S_e \left[\frac{1-F(S_e)}{1-F(1)}\right]^2$$ [equation 9]

$$F(S_e) = (1 - S_e^{*1/m})^m$$ [equation 10]

$$S_e^* = \frac{\theta_s - \theta_r}{\theta_m - \theta_r} S_e$$ [equation 11]

2_4_2 MVG ACCORDING TO VOGEL et al (2000)

In Vogel et al (2000) further changes to the formulation presented in Vogel and Cislerova (1989) are presented. A minimum capillary height h_s is adopted in order to introduce a break in the function at a certain negative value of pressure head. For this reason h_s has the same purpose of the air-entry value introduced by Brooks and Corey (1964). h_s has value ranging from -1 to -2 cm in order to keep the S-shape of VG model. The adoption of the parameter h_s is due to the necessity to have a non-zero minimum capillary height for the largest pore of the soil because it has little effects on the SWRC but large effects on the predicted shape of K(h) near saturation (Vogel et al, 2000).The parameter θ_r is substituted by the parameter θ_a with no physical meaning ($\theta_a \leq \theta_r$). The hydraulic conductivity K_k is introduced, with $K_k = K(\theta_k)$ and $\theta_k \leq \theta_s$ and $K_k < K_s$. The Modified van Genuchten model proposed by Vogel et al (2000) could be written as follows [see equations from 12 to 16]:

$$\theta(h) = \begin{cases} \theta_a + \frac{\theta_m - \theta_a}{(1+|\alpha h|^n)^m} & h < h_s \\ \theta_s & h \geq h_s \end{cases}$$ [equation 12]

$$K_r = \frac{K_k}{K_s}\left(\frac{S_e}{S_{ek}}\right)^{1/2}\left[\frac{F(\theta_r) - F(\theta)}{F(\theta_r) - F(\theta_k)}\right]^2$$ [equation 13]

$$K(h) = \begin{cases} Ks \; Kr(h) & h \leq h_s \\ K_k + \frac{(h-h_k)(K_s - K_k)}{h_s - h_k} & h_k < h < h_s \\ Ks & h \geq h_s \end{cases}$$ [equation 14]

$$F(\theta) = \left[1 - \left(\frac{\theta - \theta_a}{\theta_m - \theta_a}\right)^{1/m}\right]^m$$ [equation 15]

$$S_{ek} = \frac{\theta_k - \theta_r}{\theta_s - \theta_r}$$ [equation 16]

The two models presented do not take into consideration the wetting-drying hysteresis in soil hydraulic behavior. In most of the cases it is difficult obtaining laboratory data along the wetting path. The laboratory and field methods are available but more complex and time consuming with respect to the ones used to characterise the drying path (Likos et al, 2014).

2_5 VG MODEL CONSIDERING THE HYSTERETIC BEHAVIOUR

Hydraulic hysteresis refers to the non-unique relationship between water content and soil water potential: an important variation in the water content can occur under the same water potential depending on the history of drying/wetting to which the sample has been subjected.

In order to consider the hysteretic behaviour, both the main drying and wetting curves have to be known. The scanning curves are obtained scaling the main drying-wetting curves using different scaling approach (the slope method, the point method, the domain method etc..). All the methods have almost the same accuracy with which the scanning curves are simulated.

Hysteretic loops are represented assigning separate parameters: $\theta_s^d\ \theta_r^d\ \theta_m^d\ \alpha^d\ n^d$ to the drying path ; $\theta_s^w\ \theta_r^w\ \theta_m^w\ \alpha^w\ n^w$ to the wetting path. To reduce the number of parameters the following constraints could be adopted (see equation 17):

$\theta_r^d = \theta_r^w$; $\alpha^d \leq \alpha^w$; $n^d = n^w$ [equation 17]

Additional constraints could be used in case of lack of laboratory data (see equations 18 and 19):

$\alpha^w = 2\ \alpha^d$ [equation 18]

and

$$\theta_m^w = \theta_r + \frac{\theta_s^w - \theta_r}{\theta_s^d - \theta_r}(\theta_m^d - \theta_r)$$ [equation 19]

These relations find laboratory confirmation in the works of Simunek et al (2006); Kool and Parker (1987); Nielsen and Luckner (1992), Pham et al (2005) among others. The use of the suggested constrains allows to incorporate the hysteretic effect in the numerical flow model without increase the effort and with few laboratory data required. Moreover the simulations including hysteretic behaviour show a higher agreement with the experimental data (Kool and Parker, 1987).

Using the SWRC model elaborated by van Genuchten (1980), the Initial Drying Curve (IDC) and the Main Wetting Curve (MWC) are described as follows according to Kool and Parker (1987) model [see equations 20 and 21].

$$S_e = \frac{\theta - \theta_r^d}{\theta_s^d - \theta_r^d} = [1 + (\alpha^d \Psi^d)^{n^d}]^{-m^d} \qquad \text{(IDC)}$$ [equation 20]

$$S_e = \frac{\theta - \theta_r^w}{\theta_s^w - \theta_r^w} = [1 + (\alpha^w \Psi^w)^{n^w}]^{-m^w} \qquad \text{(MWC)}$$ [equation 21]

Where Ψ is the soil suction (in kPa). All the other parameters have been already presented. The empirical model elaborated by Scott et al (1983) is then introduced and applied. The water content on a DSC (dry scanning curve) at soil suction Ψ is scaled from IDC that is considered the reference curve, when the reversal point (θ_Δ , Ψ_Δ) is given (Simunek et al, 2012). The reversal point is the point when the drying and wetting processes reverse.

$$\theta(\Psi) = \theta_r^{ds} + \beta[\theta^d(\Psi) - \theta_r^d]$$ [equation 22]

$$\beta = \frac{(\theta_\Delta - \theta_r^d)}{(\theta^d(\Psi) - \theta_r^d)}$$ [equation 23]

$\theta_r^{ds} = \theta_s - \beta\,(\theta_s^{\,d} - \theta_r^{\,d})$ [equation 24]

$$\theta_s = \theta_s^d - \frac{(\theta_s^d - \theta_\Delta)}{(1 + R(\theta_s^d - \theta_\Delta))}$$ [equation 25]

$$R = \frac{1}{(\theta_s^d - \theta_s^w)} - \frac{1}{(\theta_s^d - \theta_r^d)}$$ [equation 26]

Where $\theta^d(\Psi)$ in equation 22 is the volumetric water content (vwc) on the IDC at soil suction Ψ; β is a parameter related to the reversal point; θ_s in equation 25 is the volumetric water content for a scanning curve at zero suction; θ_r^{ds} is the residual volumetric water content for a drying scanning curve. In the same way, for the MWC the following equations (from 27 to 29) are obtained, when the reversal point is given:

$$\theta(\Psi) = \theta_r^{ws} + \beta[\theta^w(\Psi) - \theta_r^w]$$ [equation 27]

$\theta_r^{\,ws} = \theta_s - \gamma\,(\theta_s^{\,w} - \theta_r^{\,w})$ [equation 28]

$$\gamma = \frac{(\theta_\Delta - \theta_s)}{(\theta^w(\Psi) - \theta_s^w)}$$ [equation 29]

where $\theta_r^{\,ws}$ in equation 27 is the residual vwc for the wetting scanning curve, $\theta^w(\Psi)$ in equation 27 is the volumetric water content on the MWC at the suction Ψ and γ is a parameter related to the reversal point. To reduce the number of parameters in the hysteretic model the following constrains are adopted:

$\theta_r^d = \theta_r^w = \theta_r$ [equation 30]

$\theta_s^d = \theta_s^w$ (valid if the air entrapment in the wetting process is negligible) [equation 31]

$n^d = n^w = n$ (constrain used to define a K(h) non-hysteretic) [equation 32]

Under the constraints (equation 30) and (equation 31) the eight parameters of the Kool and Parker (1987) model ($\theta_s^{\,d}\ \theta_s^{\,w}\ \theta_r^{\,d}\ \theta_r^{\,w}\ \alpha^d\ \alpha^w\ n^d\ n^w$) reduce to six ($\theta_s^{\,d}\ \theta_s^{\,w}\ \theta_r\ \alpha^d\ \alpha^w\ n$), that can reduce to five using also constrains of equation 32 ($\theta_s\ \theta_r\ \alpha^d\ \alpha^w\ n$).The last constrain (equation 30) is the strongest because it weakens the flexibility of the model in representing the water retention behaviour (Liu et al, 2018; Kool and Parker, 1987).

An analogous procedure is applied to the unsaturated hydraulic conductivity K(h) for which, in the hysteresis loop, the two functions $K^d(h)$ and $K^w(h)$ are characterised by the parameters of the SWRC in the drying and wetting branches and the saturated conductivity (Ks^d and Ks^w) (Simunek et al, 2016).

2_6 SEEPAGE ANALYSIS

Seepage analysis through porous media is of extreme interest in engineering disciplines for all that concerns design of retention structures for hydric bodies (river, lake, sea), control of contaminants in the ground and in the aquifer, slope stability analysis. Of extreme importance for a reliable stability analysis assessment of embankment sections is a proper definition of pore water pressure and water content distributions that have an important influence not only on the hydraulic response of the bank but also on

its mechanical behaviour. At the same way the transient seepage analysis through a dam subjected to fluctuation of the reservoir level is strongly affected by the conditions in the unsaturated zone that have to be taken in account properly (Freeze, 1971; Lam and Fredlund, 1984). Water is free to flow through the interconnected pores under a hydraulic gradient and the phenomenon is described using Bernouilli's theorem (gravitational head and pore water pressure head are the main driving forces in geotechnical problems) and Darcy's law. Richards (1931) for the first time understood that Darcy's low, which was originally elaborated for saturated soil, could be equally applied to unsaturated soil and the permeability is not a constant but function of the matric suction of the soil, K(h). The pressure gradient that drives the gaseous phase is neglected in Richards' equation. Richards' equation could be written in a number of forms: in 1D the mixed water content form (because it mixes water content and water potential), in water content form (only the water content is present) or in water potential form (only water potential is present) (Farthing and Ogden, 2017). Here in equation 33 the water potential based form is presented:

$$\frac{d\theta}{dt} = \frac{d\theta}{d\Psi}\frac{d\Psi}{dt} = C(\Psi)\frac{d\Psi}{dt} \qquad \text{[equation 33]}$$

where θ is the water content, Ψ is the water potential and C is the specific water capacity [L^{-1}] that is the slope of the SWRC. The Ψ-based form is applicable to unsaturated and near-saturation conditions, but the mass balance error could be relevant and the convergence could slow (Hendriks, 2010).

In 2D, Richard equation could be written as follows in equation 34 (neglecting the water loss through root uptake and evaporation) (Simunek et al,2006).

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial x_i}\left[K\left(K_{ij}^{A}\frac{\partial\Psi}{\partial_{X_j}}\right) + K_{iz}^{A}\right] \qquad \text{[equation 34]}$$

$$K(\Psi, x) = K_{sat}\ (x, z)\ K_r\ (h, x, z) \qquad \text{[equation 35]}$$

where x_i are the spatial coordinates (i=1,2), Kij ^{A} and Kiz ^{A} are component of the anisotropy tensor K^A . If the flow is isotropic (i.e. the hydraulic permeability K is equal in the horizontal and vertical direction) K^A has element equal to 1 in the diagonal and equal to 0 in the off-diagonal.

Richards' equation is a highly non-linear partial differential equation, almost impossible to solve in a closed form with the exception of a small number of cases (Miller et al, 1998). This is due to the fact that the relation is dependent on two highly non-linear equations: the hydraulic conductivity function K (θ) and the water potential Ψ (θ). When saturation is close to 100% Ψ is a value near zero, for big and non-positive value of Ψ, K is close to zero. This extreme cases lead to degeneracy in the equation (Farthing and Ogden, 2017).

The first attempt to solve seepage problems was graphical, using the so called ''flow-net method'', proposed by Casagrande (1937). The method could be applied for simple problems in which soil is isotropic and homogeneous and the hypothesis at the base of the method is acceptable (water flows only in the saturated zone). The first attempt to model by finite elements the flow in the saturated zone was made by Taylor and Brown (1967) in which the water table is assumed as upper boundary condition and using a trial and error procedure its position is located in the model. Freeze (1971) proposed for the first time a ''saturated-unsaturated model'' using a finite difference model in which pressure heads are calculated also for the unsaturated zone (Lam and Fredlund, 1984; Lam and Fredlund, 1993). With the innovation brought

by high-speed digital computers, numerical methods, especially finite element methods, are now able to give solution to complex seepage problems. The lack of a solution, reliability, convergence problems (slow or not possible) are yet unsolved problems that impose a penalty on the use of Richard's equation solvers (Farthing and Ogden, 2017). A robust and accurate solution methodology to be applied to a large variety of soil has not yet been identified.

2_7 NUMERICAL DIRECT AND INVERSE MODELLING USING HYDRUS SOFTWARE

Hydrus program is a computer package to simulate water flow, pollution or heat propagation through a saturated or unsaturated porous media, under steady state or variable boundary conditions. This is done in a 1D space using Hydrus 1D, in a 2D (vertical and horizontal planes) and 3D variably saturated domains (axisimmetrical or fully 3D) using Hydrus 2D/3D.

Hydrus is a powerful tool that traces its origin from the innovative work and first computer software elaboration of van Genuchten at the end of the seventy's (SUMATRA and WORM), of Vogel at the end of the eighty's (SWMI), later of Kool and van Genuchten (Hydrus) and Simunek in 1998 (Hydrus 1D) (Simunek et al, 2012). Hydrus 1D, 2D and 3D have become widely used software, whose codes were reviewed favourably several times and on whose application to real case studies a wide bibliography is available (Simunek et al, 2012).

The program solves Richards' equation for saturated and unsaturated flows using mass-lumped linear finite element schemes in direct problems, when model parameters, initial and boundary conditions are known. Hydrus allows running also inverse problem in which one or more parameters are calibrated or estimated from observed data. Model calibration and inverse parameters estimation are performed using an implemented local optimization approach based on Levenberg-Marquardt method which will be discussed in details in the following chapter §2_7_1.

Different models have been implemented in Hydrus code to describe unsaturated soil hydraulic behaviour: van Genuchten (1980), Brooks and Corey (1964), modified van Genuchten (Vogel and Císlerová, 1988), Durner (1994), and Kosugi (1996). Soil hysteretic behaviour could be incorporated in the van Genuchten model (1980) as seen in chapter §2_5. Hydrus code can simulate a great number of processes such as precipitation, infiltration, irrigation, evaporation, capillary rise, root water uptake, soil water storage. Hydrus has been applied successfully to a wide range of spatial and temporal scales: from the 1D laboratory experiment to the modelling of transport domain hundreds of meters wide. Due to the highly non-linearity of Richards' equation it is preferable to maintain a medium scale because the spatial discretisation of the geometry requires a fine-mesh close to atmospheric boundaries where large hydraulic gradients are expected. This leads to convergence problems or very long calculation time (Simunek et al, 2012).

As extensively discussed in chapter §2_6, Richards' equation describes water flow in an unsaturated porous media. Hydrus solves the equation implementing the soil hydraulic functions of VG-Mualem to describe unsaturated conductivity in terms of water retention parameters (Simunek et al, 2006). Hydrus uses a Finite Element numerical approach to solve Richards' equation in space and a Finite Difference approach in time. The finite difference approach uses the mixed form of the Richard equation to develop a set of algebraic equations.

2_7_1 MODEL CALIBRATION AND INVERSE ESTIMATION OF THE HYDRAULIC PARAMETERS

Hydrus software could be used for direct problems (model parameters, boundary conditions and initial conditions are known) and indirect problems (one or more parameters are unknown and need to be estimated from observed data or some parameters values have an high degree of uncertainty and they need to be calibrated properly in the model).

The models, which are most often used, are parametric models i.e. they include parameters. Parameters could be defined as numeric factors in the model equations that could make the model flexible assuming different values. As their value change, the flexible shape of the equations changes while the structure is maintained. Parameters could be of different nature and they could assume a fixed value, a value dependent on time or on the state of the system. Parameters adapt and optimize model performances while model tries to represent natural phenomena. To be applied to a model, all parameters need to be initially estimated, then calibrated, then optimized and after that, we can state that the model is calibrated, parameterized and optimized.

Model calibration is the procedure to tune a model for a specific problem by manipulating the input parameters (e.g. soil hydraulic parameters), boundary conditions, initial conditions in a reasonably range of values until the output of the model closely match the observed values (e.g. water content, pressure head, fluxes, concentrations etc) (Simunek and Hopman, 2002).

The most common and used approach in model calibration is the selection of an objective function that measures the agreement between predicted data (by the model) and the measured data. The minimization of the objective function gives in output the best-fit parameters. In the past, calibration was performed by a trial and error technique, a very time consuming procedure that does not ensure good results. The calibration procedure is concluded when the model is able to reproduce data within a subjectively acceptable level of precision (Konikow and Bredehoeft, 1992). When the main goal of the calibration procedure is not to obtain a better fit between measured and observed data but rather to optimize unknown parameters in that model, the process takes the name of ‘’*parameter optimization*’’ or ‘’*parameter estimation*’’ (Simunek and de Vos, 1999).

It is important at this point to make a summary of the various routinely analysis used in the theory and application of the inverse modelling. The optimisation of the parameters is performed by systematically minimizing the sum of the squares of the differences between observed and simulated variables. This sum is expressed by an objective function ϕ (β) [equation 36]:

$$\phi(\beta, q, p) = \sum_{j=1}^{m_q} v_j \sum_{i=1}^{n_{q\,j}} w_{i,j}[q_j^*(x, t_i) - q_j(x, t, \beta)]^2 \qquad \text{[equation 36]}$$

where q* are the measured variables, q the estimated variables by model predictions, β is the vector of the optimized parameters describing, in our case, unsaturated soil hydraulic properties (i.e. soil water retention parameters, hydraulic permeability function). m_q in the first summation represents the different measurement types (i.e. pressure head, water content), n_q is the number of measurements for a certain type j of measurement. v_j is the weight used to homogenize different data types by means of a normalization procedure or assuming v_j equal to the measurement variance of measurement type j, w_{ij} are the weights assigned to the individual data (Hopmans et al 2002;.Simunek et al, 2002).

ϕ (β) is called also ‘’*weighted least squares estimator*’’ and as long as the measurement errors are uncorrelated and independent (i.e. the weight w_{ij} contains the measurement error information), it is also the maximum-likelihood estimator (it maximizes the likelihood function).

$$w_i = \frac{1}{6_i^2} \qquad \text{[equation 37]}$$

where 6_i is the variance of the measurement error of q_i*.

The objective function implemented in Hydrus adds to the weighted least square estimator additional information and it is defined as the sum of three components:

$$\phi(\beta, q, p) = \sum_{j=1}^{m_q} v_j \sum_{i=1}^{n_{q,j}} w_{i,j}[q_j^*(x,t_i) - q_j(x,t_i,\beta)]^2 + \sum_{j=1}^{m_p} v_j \sum_{i=1}^{n_{p,j}} w_{i,j}[p_j^*(\theta_i) - p_j(\theta_i,\beta)]^2 + \sum_{j=1}^{n_b} \hat{v}_j\,[\beta_j^* - \beta_j]^2$$

[equation 38]

Where the first term has been already presented and discussed, the second represents the differences between independently measured and predicted soil hydraulic parameters for all the investigated layers. p*, p, $\overline{w_{1J}}$, $\overline{v_J}$, m_p and n_p have a similar meanings as the first term. The third term is a penalisation term between initial estimate (β_j*) of the hydraulic parameters and their final estimate β_j ,n_b is the number of parameters with prior knowledge and $\hat{v_j}$ are the pre-assigned weights.

Hydrus software package has implemented a Levenberg-Marquardt method (Marquardt, 1963; Simunek and Hopmans, 2002) to minimize the objective function in order to perform inverse estimation of hydraulic, solute and heat transport parameters in steady-state or transient flow models. Levenberg-Marquardt method has become a standard in nonlinear squares fitting, it combines the Newton method with the steepest descend method. This method assumes that the covariance matrices are diagonal. The covariance matrix gives information on possible correlation between parameters and on the measurement accuracy (Simunek and Hopman, 2002). The most significant advantage of Levenberg- Marquardt method is its computational efficiency because it needs far fewer direct model evaluations compared to others resolution techniques (Simunek et al, 2012). The Levenberg-Marquardt method needs an initial estimate of the unknown parameters to be optimized. The objective function ϕ is computed in the neighborhood of the initial estimate to understand which direction has to be taken in order to optimize the parameters.

2_7_2 MODEL PERFORMANCE EVALUATION BY STATISTICAL ANALYSIS

The most frequently used calibration procedure is through the optimization of model performances which is carried out comparing simulated-observed data. This is the procedure to which particular attention will be paid in this chapter and in the followings of the present thesis.

Levels of acceptable accuracy and precision are set to evaluate if a model gives a good representation of an observable natural phenomenon. No model could be detailed enough to be valid for all the simulations, this is the reason why the modeller must decide which processes have to be modelled and the level of desired accuracy and precision to be set to validate the model.

The typical approach adopted to evaluate **model performance** uses the comparison between simulated output and a set of observations. Model performance can be addressed by means of qualitative and

quantitative criteria. Qualitative criteria essentially rely on the graphical comparison between observed and simulated data, whereas quantitative criteria are based on numerical performance metrics (or indices). Both approaches are fundamental tools to be used. The choice of the validation criteria is guided by several factors such as the nature of the simulated variables, the main model purposes and the simulation model itself. There are many metrics able to take into consideration the different sources of uncertainty associated with input data, model structure and parameterization. Many remarks have been addressed to traditional lumped metrics for their lack of diagnostic power or inability to capture differences between different model or parameter sets leading to ambiguous evaluations.

We present a review of robust performance metrics (or indices) to provide a quantitative estimate of model reliability. These metrics could be divided in three main groups.

- The *Accuracy metrics* are able to evaluate the ability of the model to reproduce the observed phenomenon; the accuracy is largely dependent on the accuracy of the model input parameters and on the model ability (by means of the implemented equations) to represent successfully the physical system.
- The *Efficiency metrics* are used to assess the quality of a model performance measuring the deviance between estimated and true values and the variance of the two input datasets is the key parameter to guide the analysis.
- The *Model selection metrics* are used commonly for model evaluation and selection. These metrics take into account the goodness of the fit and the model complexity in terms of number of parameters. These indices are based on the entropy concept since they are able to measure the quantity of the dataset information that are lost when a model is used to represent a natural phenomenon. This type of metrics is extremely relevant because the maximum-likelihood approach adopted in the inverse problem performs the optimization of the parameters without questioning the adequacy of the model used.

All these indices will guide the calibration phase of the investigated model that is divided in one pre-calibration step and three main steps. The pre-calibration step is a phase of skimming of the large number of performed inverse analysis using a small selection of metrics. The first, second and third steps lead to the identification of the sets of optimized parameters using every time different groups of statistical metrics that highlight different aspects of the inverse problem.

2_7_2_1 ACCURACY METRICS

Common metrics to evaluate the forecast accuracy include MAE (Mean Absolute Error), MBE (Mean Bias Error), MSE (Mean Square Error), RMSE (Root Mean Square Error) ,CRM (Coefficient of Residual Mass), SEC (Standard Error Coefficient) and NCOV (Normalized Coefficient of Variation). It is always important to remember that MAE, MBE and RMSE have the same units of the variable under investigation and thus cannot be compared to variables that are scaled differently. In order to eliminate the influence of the different dimensions in the accuracy metrics MSE, MBE, MSE, RMSE, metrics are divided by the mean of the observed values ($\overline{M}$) as proposed by Cai et al (2017). Accuracy metrics could be used in the direct problem to evaluate the accuracy of a model and to guide the calibration process in the indirect problem but also to guide the evaluation of the performance in the validation process on newly observed-simulated dataset (cross-validation).

- MAE (Mean Absolute Error)

$$\mathrm{MAE} = \frac{1}{\mathrm{n}} \cdot \sum_1^{\mathrm{n}} |\mathrm{M_i} - \mathrm{S_i}| \qquad \text{[equation 39]}$$

Where M_i is the observation value, S_i is the simulated value and ''n'' the number of available real data. MAE is the simplest measure of forecast accuracy and it is simply, as the name suggests, the mean of the absolute values of the difference between the real value and the forecasted value, where all individual differences have the same weight. It considers the average magnitude of the errors in a set of predictions without considering their direction (positive or negative). MAE can range from 0 (best fit) to infinite and it expresses the average model prediction error in units of the variable of interest.

- MBE (Mean Bias Error)

$$\mathrm{MBE} = \frac{1}{\mathrm{n}} \cdot \sum_1^{\mathrm{n}} (\mathrm{M_i} - \mathrm{S_i}) \qquad \text{[equation 40]}$$

Where M_i is the observation value, S_i is the simulated value and ''n'' the number of available real data. The Mean Bias Error is usually not used as a measure of the model error because high individual errors in the prediction could also give a low MBE. Mean Bias Error is primarily used to estimate the average bias in the model and to decide if any steps need to be taken to correct the model bias. *MBE* can range from 0 (best fit) to infinite and it captures the average bias in the prediction.

- MSE (Mean Square Error)

$$\mathrm{MSE} = \frac{1}{\mathrm{n}} \cdot \sum_1^{\mathrm{n}} (\mathrm{M_i} - \mathrm{S_i})^2 \qquad \text{[equation 41]}$$

Where M_i is the observation value, S_i is the simulated value and ''n'' the number of available real data. The MSE is calculated as the average square difference between simulated and observed data and it is always strictly positive. MSE has the same units of the square of the estimated quantity in the brackets. MSE could range from 0 (best fit) to infinite.

- RMSE (Root Mean Square Error)

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}} \cdot \sum_1^{\mathrm{n}} (\mathrm{M_i} - \mathrm{S_i})^2} \qquad \text{[equation 42]}$$

Where M_i is the observation value, S_i is the simulated value and ''n'' the number of available real data. RMSE, called also Root Mean Square Deviation (RMSD), is a worldwide used parameter to measure the difference between values predicted by model and values observed. The individual differences (in brackets) are called ''*residuals*''.

An alternative formulation of RMSE that can be used when different weights are given inside a specific dataset of measures is the following (Simunek and Hopmans, 2002) :

$$\mathrm{RMSE} = \sqrt{\frac{1}{(\mathrm{n}-\mathrm{m})} \cdot \sum_1^{\mathrm{n}} \mathrm{w_i} (\mathrm{M_i} - \mathrm{S_i})^2} \qquad \text{[equation 43]}$$

Where $''\mathrm{n}''$ is the number of recorded measurements and $''\mathrm{m}''$ is the number of estimated parameters, (n-m) is the degree of freedom and w_i the weight associated to each measurement of a dataset.

Loague and Green (1991) proposed to divide RSME by the mean of the observed values. However, relate RMSE to the average of the observed data produces an unstable index with values very close to zero. Despite this, the normalization of RMSE facilitates the comparison between datasets or models that use different scales. For this reason, Loague and Green introduced another very useful index, CRM, the coefficient of residual mass (Loague and Green, 1991).

- CRM (Coefficient of Residual Mass)

$$\mathrm{CRM} = \frac{(\sum_1^n \mathrm{M_i} - \sum_1^n \mathrm{S_i})}{\sum_1^n \mathrm{M_i}} \qquad \text{[equation 44]}$$

Where M_i is the observation value, S_i is the simulated value and ''n'' the number of available real data. CRM measures the model tendency to underestimate or overestimate the measurements. The model underestimates the measurements when CRM has a positive value, while the model has a tendency to overestimate when CRM has a negative value. As the model efficiency NSE, CRM has been introduced by Loague and Green (1991). For a perfect model, CRM is equal to 0 (best fit).

If the CRM is expressed in percentage, it assumes the name of ''*Deviation of runoff Volume*'', introduced by the World Meteorological Organization WMO (1986).

$$\mathrm{Dv}\ (\%) = \frac{(\sum_1^n M_i - \sum_1^n S_i)}{\sum_1^n M_i} * 100 \qquad \text{[equation 45]}$$

- SEC (Standard Error Coefficient/Standard Error Parameter)

In statistics variance σ_{X}^2 measures how far a set of data $\mathrm{x_1}, \ldots\ldots, \mathrm{x_n}$ is distributed with respect to its mean value $\bar{\mathrm{x}}$. The variance is defined as follow [equations 46, 47,48]

$$\sigma_X^2 = \frac{\sum_1^n (x_i - \bar{x})^2}{n} \qquad \text{[equation 46]}$$

$$X = x_1, \ldots\ldots, x_n \qquad \text{[equation 47]}$$

$$\bar{\mathrm{x}} = \frac{\sum_1^n \mathrm{x_i}}{\mathrm{n}} \qquad \text{[equation 48]}$$

The variance is the square of the standard deviation σ_{X}.

$$\sigma_X = \sqrt{\frac{\sum_1^n (x_i - \bar{x})^2}{n}} \qquad \text{[equation 49]}$$

The standard deviation is a measure of the amount of variation (dispersion) of a set of data. A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) while a high standard deviation indicates that the values are distributed over a wider range. Standard deviation is a measure of uncertainty. The standard deviation of a group of repeated measurements gives information on the precision of these measurements. If the modeler needs to understand if measurements agree or not with a theoretical prediction, the standard deviation of these measurements is of crucial importance. If the mean of the measurements is too far from the prediction (with the distance measured using the standard deviation), then the theory being tested probably needs to be revised. This makes sense since the measurements fall outside the range of values that could reasonably be expected to occur if the predictions were correct and the standard deviation was appropriately quantified. Standard deviation could be used to

obtain also confidence intervals on the variance of residuals from a least squares fit under standard normal theory, where ”n” is now the number of degrees of freedom (see chapter §2_7_2_5).

Now we can define the standard error coefficient (SEC). Standard error is a statistical index that measures the accuracy with which a sample distribution represents a population by using standard deviation. Usually SEC refers to the standard deviation of a statistical parameter such as the mean or the median. For example the standard error of the mean $\mathrm{SEC}(\bar{x})$ expresses the standard deviation of the distribution of sample means sampled from a population. The general formulation is:

$$\mathrm{SEC}(\bar{x}) = \frac{6}{\sqrt{n}} \qquad \text{[equation 50]}$$

Where 6 is the sample standard deviation and ‘’n’’ the sample size.

$\mathrm{SEC}(\bar{x})$ is calculated taking the standard deviation and dividing it by the square root of the sample size. It depends on both the standard deviation and the sample size. It is inversely proportional to the sample size: the larger the sample size, the smaller the standard error. If we want to investigate how widely scattered some measurements are, we can use the standard deviation. If we want to investigate the uncertainty around the estimate of the mean measurement, we can use the standard error of the mean.

The concept of the Standard Error Coefficient can be applied to a system of equations for which the significance of each parameter aims to be investigated. The application to the system of flow equations solved by the Levenberg-Marquardt (LM) algorithm is presented.

In the following, vectors and arrays appear in bold. Let ‘’ f ‘’ be an assumed functional relation which maps a parameters vector $\boldsymbol{\theta} \in \mathbb{R}^n$ to an estimate measurements vector $\hat{\boldsymbol{x}}$ **:**

$$\boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{\theta}) = \mathbf{0} \quad \boldsymbol{f} \in \mathbb{R}^{\boldsymbol{m}} \quad \hat{\boldsymbol{x}} \in \mathbb{R}^{\boldsymbol{m}} \quad \boldsymbol{\theta} \in \mathbb{R}^n \quad m > n \qquad \text{[equation 51]}$$

$$\boldsymbol{r}(\boldsymbol{\theta}) = \boldsymbol{f}(\hat{\boldsymbol{x}}, \boldsymbol{\theta}) \quad \text{(residuals)} \qquad \text{[equation 52]}$$

$$r_i = f_i(\hat{\boldsymbol{x}}, \boldsymbol{\theta}) \quad i = 1, m \qquad \text{[equation 53]}$$

The LM algorithm minimizes the objective function **Φ** :

$$\Phi(\boldsymbol{\theta}) = \sum_1^m r_i^2(\boldsymbol{\theta}) \qquad \text{[equation 54]}$$

Starting with an initial guess $\boldsymbol{\theta}^{(\mathbf{0})}$ for the minimum, the method proceeds by iterations:

$$\boldsymbol{\theta}^{(\boldsymbol{k}+\mathbf{1})} = \boldsymbol{\theta}^{(\boldsymbol{k})} + \Delta\boldsymbol{\theta} \qquad \text{[equation 55]}$$

$$\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right) \cdot \Delta\boldsymbol{\theta} + \boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{r} = \mathbf{0} \qquad \text{[equation 56]}$$

$$(\boldsymbol{J_{res}})_{\boldsymbol{ij}} = \frac{\boldsymbol{\partial r_i(\theta^{(k)})}}{\boldsymbol{\partial \theta_j}} \qquad \text{[equation 57]}$$

$$\Delta\boldsymbol{\theta} = -\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right)^{-1} \cdot \boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{r} \qquad \text{[equation 58]}$$

where $\boldsymbol{J_{res}}^{\boldsymbol{T}}$ denotes the matrix transpose.

The matrix $\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right)$ on the left hand side is the approximate Hessian.

The residual standard deviation is defined as:

$$\sigma_{res} = \sqrt{\frac{\sum r_i^2}{m-n}} \qquad \text{[equation 59]}$$

$$df = degree\ of\ freedom = m - n$$

It is relevant to note the likeness between this formula and the general formulation SEC (degree of freedom instead of sample size, sum of quadratic residuals instead of the standard deviation).

To calculate the proper standard error for a given parameter the following formula is applied:

$$\sigma_{\theta_j} = \sigma_{res} \cdot \sqrt{H_{jj}} \qquad \text{[equation 60]}$$

Where H_{jj} is the j_th diagonal element of the inverse of the Hessian matrix:

$$\left({J_{res}}^T \cdot J_{res}\right)^{-1} \qquad \text{[equation 61]}$$

σ_{θ_j} provides a measure of how precisely each parameter has been estimated. It must, however, be interpreted with care. In fact its computation is based on the assumption that the data points are statistically independent. Under such circumstances the Parameter Standard Error underestimates the true error.

We are often interested in verifying hypotheses on one single estimated parameter in order to understand if the model under investigation is parameter redundant.

A t-test could be used to evaluate the significance of each parameter:

$$H_0\ hypothesis \rightarrow \quad \theta_J = 0$$

$$H_1\ hypothesis \rightarrow \quad \theta_J \neq 0$$

If the $\mathrm{H_0}$ hypothesis is true, θ_J parameter could be eliminated from the model. The t-value is equal to:

$$t_0 = \frac{\theta_J}{\sigma_{\theta_j}} \qquad \text{[equation 62]}$$

Once the statistic value t_0 is determined, we have to read in t-test table the critical value $\mathrm{t_{\alpha,m-n}}$ of the Student's t-distribution described by $''df = m - n\,''$ degree of freedom corresponding to the significance level α of our choice (5% i.e. $\alpha = 0.05$). If $t_0 \geq t_{\alpha,m-n}$ we reject $\mathrm{H_0}$ that means that the parameter θ_J is significant.

- NCV (Normalized Coefficient of Variation)

The Coefficient of Variation (CV), also known as the Relative Standard Deviation, is a statistical measure of the dispersion of a data set around the mean. It is calculated as the ratio of the standard deviation to the mean and it is usually expressed in percentage (NCV).

$$\mathrm{CV} = \frac{\sigma_X}{\bar{x}} \qquad \text{[equation 63]}$$

$$\mathrm{NCV} = \left(\frac{\sigma_X}{\bar{x}} * 100\right)$$ [equation 64]

where: $X = x_1, \ldots\ldots, x_n$ and $\bar{\mathrm{x}} = \frac{\sum_1^{\mathrm{n}} \mathrm{x_i}}{\mathrm{n}}$.

The Coefficient of Variation (CV) is sometimes preferred to the standard deviation because it is independent of the unit of measurement. When comparing the variability between datasets with different measurement scales or very different mean values, the coefficient of variation can be a useful alternative or complement to the standard deviation. Higher the CV, greater the spread of the data around the mean. However, the coefficient of variation should not be used when the mean value is close to zero. The coefficient of variation can approach infinity and its value is susceptible to small changes in the mean.

The downside is that CV is invariant to the number of data so it does not show the increase in the measurements certainty with the increasing number of input data. For this particular limit, the standard deviation is a superior indicator.

2_7_2_2 EFFICIENCY CRITERIA

In this section, the Efficiency Criteria used in this study are presented and discussed. In particular five criteria have been used:

- Coefficient of determination.
- Nash-Sutcliffe efficiency.
- Nash- Sutcliffe efficiency with logarithmic values.
- Index of Agreement, together with four modified forms that may provide more information on the systematic and dynamic errors present in the model simulation.

Below a series of definitions, used later to introduce efficiency criteria parameters, are reported:

$\bar{X} = \frac{1}{n} \cdot \sum_1^n X_i$ (mean value) [equation 65]

$\sigma_X^2 = \frac{1}{n} \cdot \sum_1^n (X_i - \bar{X})^2$ (variance) [equation 66]

$\sigma_X = \sqrt{\frac{1}{n} \cdot \sum_1^n (X_i - \bar{X})^2}$ (standard deviation) [equation 67]

$Cov\ (X, Y) = \frac{1}{n} \cdot \sum_1^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})$ (covariance) [equation 68]

- NSE (Nash-Sutcliffe Efficiency)

$$NSE = 1 - \frac{\sum_1^n (M_i - S_i)^2}{\sum_1^n (M_i - \bar{M})^2}$$ [equation 69]

Where M_i is the observation value, S_i is the simulated value, ''n'' the number of observed data and $\bar{M}$ the mean of the observed values. It is used in almost every field of simulation. The NSE criterion is a normalized statistic parameter that measures the relative magnitude of the model error variance compared to the measured data variance (Nash and Sutcliffe, 1970). The biggest disadvantage of NSE criterion is the fact that the differences observed-simulated data are calculated as squared values. In this way larger values in a time series are strongly overestimated while lower values are neglected (Legates and McCabe, 1999)

NSE = 1 corresponds to a perfect match between modelled and observed data;

NSE = 0 indicates that the model predictions are as accurate as the mean of the observed data;

NSE < 0 indicates that the mean of the observed data is a better predictor than the model.

NSE is born and firstly used to calibrate parameters of hydrological models, to compare modelling results and to communicate the results to stakeholders or to other modellers (Moussa et al, 2010).NSE in flow models is often applied to short time period (a day or couple of days) to evaluate model ability to represent extreme values (for example flow peaks). If applied to long time period (month), it is used to evaluate the general pattern of flows without considering the single events. In this case, it is not recommended to use NSE parameter to evaluate the model during the low flows because the parameter is dominated by the high flows errors (Thirel et al, 2014). This is because NSE is not sensitive to systematic model over and under predictions especially in long-time simulations. NSE has a range from 1 (best fit) to – infinite (usually when NSE <0, the simulation is unacceptable).

- NSE log (Nash-Sutcliffe efficiency with logarithmic values)

$$NSE_{log} = 1 - \frac{\sum_1^n \left(M_i{}^{log} - S_i{}^{log}\right)^2}{\sum_1^n \left(M_i{}^{log} - \overline{M^{log}}\right)^2} \quad \text{[equation 70]}$$

$$M_i{}^{log} = log(M_i + c) \quad \text{[equation 71]}$$

$$S_i{}^{log} = log(S_i + c) \quad \text{[equation 72]}$$

Where M_i is the observation value, S_i is the simulated value, ''n'' the number of observed data and $\overline{M^{log}}$ is the mean of the $M_i{}^{log}$ values with i varying from 1 to ''n''. ''c'' is a positive constant equal to the 10th percentile of the observed data. The use of this constant de-emphasises all the very small values that tend to be unrealistic and avoids numerical problem when attempting to calculate the logarithm of 0. For the same reason it is not possible to apply NSE_{log} to negative or zero because the logarithm in these cases is undefined.

To overcome the problem related to NSE sensitivity to extreme values and the squared differences in the field of hydraulic modelling, NSE has been calculated using the logarithm of the data. In this way, when NSE_{log} is applied to model of flux the flow peaks are flattened and the low flow kept at the same level. In this way the parameter is more sensible to low flow values and in general the sensitivity to systematic model over and under prediction is increased (Seiller et al, 2012).

- R Pearson's correlation coefficient

$$R = \frac{\sum_1^n (M_i - \bar{M})(S_i - \bar{S})}{\sqrt{\sum (M_i - \bar{M})^2} \cdot \sqrt{\sum (S_i - \bar{S})^2}} \quad \text{[equation 73]}$$

Where M_i is the observation values, S_i is the simulated values, ''n'' the number of observed data, $\bar{M}$ the mean of the observed values and $\bar{S}$ the mean of the simulated values.

One of most used parameter to express a measure of the dependence between two quantities is the *Pearson Product-Moment Correlation Coefficient*, called simply ''*Pearson's Correlation Coefficient*'' or ''*the Correlation Coefficient*''. It is based on the method of the covariance and with no doubt is one of the best

methods to measure the association between variables. The Correlation Coefficient gives information on the magnitude of the correlation or association and the direction of the relationship and it gives a measure of the strength of the linear relationship between two variables.

The *coefficient of determination* R^2 is defined as the squared value of the coefficient of correlation according to Bravais-Pearson. It is calculated as [equation 74]:

$$R^2 = \left\{ \frac{\sum_1^n (M_i - \bar{M})(S_i - \bar{S})}{\sqrt{\sum (M_i - \bar{M})^2} \cdot \sqrt{\sum (S_i - \bar{S})^2}} \right\}^2 \quad \text{[equation 74]}$$

R^2 can also be expressed as the squared ratio between the covariance (CV (M,S)) and the multiplied standard deviations of the observed (σ_M) and predicted values (σ_S). It estimates the combined dispersion against the single dispersion of the observed and predicted series [equations 75 and 76]:

$$R^2 = \left\{ \frac{Cov\ (M,S)}{\sigma_M \cdot \sigma_S} \right\}^2 \quad \text{[equation 75]}$$

$$\sigma_M = \sqrt{\frac{1}{n} \cdot \sum_1^n (M_i - \bar{M})^2} \qquad \sigma_S = \sqrt{\frac{1}{n} \cdot \sum_1^n (S_i - \bar{S})^2} \quad \text{[equation 76]}$$

The Covariance suffers of the same limits of the variance: it has not the unit of measures of the input data and the same order of magnitude of the values Mi and Si. In R^2 formulation, the division by $(\sigma_M \cdot \sigma_S)$ allows to have a dimensionless parameter whose absolute value no longer depends on the order of magnitude of the values Mi and Si. R^2 ranges between 0 and 1. A value of 1 means that the dispersion of the predictions is equal to the one of the observations while a value of 0 means no correlation at all. One of the major drawbacks of the R^2 value, if considered alone, is that it takes into consideration just the dispersion. It is always suggested to use this parameter together with other statistical parameters.

An alternative formulation of R^2 that could be used when different weights are given inside a specific dataset of measures is the following in equation 77 (Simunek and Hopmans, 2002):

$$r^2 = \frac{(\sum w_i M_i S_i - (\sum w_i M_i * \sum w_i S_i)/\sum w_i)^2}{(w_i M_i^2 - \frac{(\sum w_i M_i)^2}{\sum w_i})\,(w_i S_i^2 - \frac{(\sum w_i S_i)^2}{\sum w_i})} \quad \text{[equation 77]}$$

The two formulations of the Coefficient of Determination (R^2 and r^2) are equivalent if the weights w_i are all equal to 1 [equation 78, 79, 80].

$$\sum_1^n w_i = n \quad \text{[equation 78]}$$

$$\frac{\sum w_i M_i}{\sum w_i} = \frac{\sum M_i}{n} = \bar{M} \quad \text{[equation 79]}$$

$$\frac{\sum w_i S_i}{\sum w_i} = \frac{\sum S_i}{n} = \bar{S} \quad \text{[equation 80]}$$

- IA (Index of Agreement)

$$IA = 1 - \frac{\sum_1^n (M_i - S_i)^2}{\sum_1^n (|S_i - \bar{M}| + |M_i - \bar{M}|)^2} \quad \text{[equation 81]}$$

Where M_i is the observed value, S_i is the simulated value, ''n'' the number of observed data and $\bar{M}$ the mean of the observed values. The Index of Agreement proposed by Willmott (1981) is able to overcome the insensitivity of NSE and R^2 to the differences between observed and predicted variances and mean (Legates and McCabe, 1999). It represents the ratio between mean square error and potential error and it varies between 0 and 1. A value of 1 indicates a perfect match while a value of 0 no agreement at all. IA is able to detect the additive and proportional differences in the simulated and observed means and variances. The main limit of this parameter is the oversensitivity to extreme values due to the squared differences. For this reason IA, as seen for NSE, is very sensitive to peak flows and insensitive to low flow conditions.

- Modified forms of NSE and IA

$$NSE_j = 1 - \frac{\sum_1^n |M_i - S_i|^j}{\sum_1^n |M_i - \bar{M}|^j} \quad j \in N \quad \text{[equation 82]}$$

$$IA_j = 1 - \frac{\sum_1^n |M_i - S_i|^j}{\sum_1^n (|S_i - \bar{M}| + |M_i - \bar{M}|)^j} \quad j \in N \quad \text{[equation 83]}$$

Where M_i is the observed value, S_i is the simulated value, ''n'' the number of observed data and $\bar{M}$ the mean of the observed values. To overcome the oversensitivity to extreme values of NSE and IA, these general forms of the two equations could be used. In hydrologic modelling, it has been found that for j=1 the overestimation of the flood peaks is significantly reduced resulting in a better evaluation especially of the lower values while an increase of j results in an increase of the sensitivity to high flows, for this reason it is recommended to use a value of j>1 only when high flows are of interest (Krause et al, 2005). The use of j>1 to study low values must be avoided. As seen for NSE and IA, both parameters are in the range 0-1 (1 is the best fit). Due to the fact that absolute values are preferable to squared terms to avoid that errors of outliers are over weighted when squared, the choice to adopt j=1 is commonly accepted (Legates and Mc Cobe, 1999).

- Relative efficiency criteria (NSE_{rel} and IA_{rel})

$$NSE_{rel} = 1 - \frac{\sum_1^n \left(\frac{M_i - S_i}{M_i}\right)^2}{\sum_1^n \left(\frac{M_i - \bar{M}}{\bar{M}}\right)^2} \quad \text{[equation 84]}$$

$$IA_{rel} = 1 - \frac{\sum_1^n \left(\frac{M_i - S_i}{M_i}\right)^2}{\sum_1^n \left(\frac{|S_i - \bar{M}| + |M_i - \bar{M}|}{\bar{M}}\right)^2} \quad \text{[equation 85]}$$

All criteria described above are able to quantify the differences between observed and predicted values in absolute value. The result gives an over or under prediction of the higher values while lower values have a low influence on the parameter value. For this reason the parameters NSE_{rel} and IA_{rel} based on relative deviation and derived from NSE and IA are proposed. As seen for NSE and IA, both parameters are in the range 0-1 (1 is the best fit).

- KGE (Kling-Gupta Efficiency)

Gupta et al (2009) proposed a diagnostically interesting decomposition of the Nash-Sutcliffe efficiency parameter which could facilitate the analysis of the relative importance of its different components (correlation, bias and variability) in the context of hydrological modelling. Kling et al (2012), proposed a revised version of this index, to ensure that the bias and variability ratios are not cross-correlated.

$$KGE = 1 - \sqrt{(R-1)^2 + \left(\frac{\sigma_S}{\sigma_M} - 1\right)^2 + \left(\frac{\bar{S}}{\bar{M}} - 1\right)^2} \qquad \text{[equation 86]}$$

$$\sigma_M = \sqrt{\frac{1}{n} \cdot \sum_1^n (M_i - \bar{M})^2} \qquad \sigma_S = \sqrt{\frac{1}{n} \cdot \sum_1^n (S_i - \bar{S})^2} \qquad \text{[equation 87]}$$

- Mass Balance Error

Mass Balance Error is usually displayed in the postprocessor. It gives an indication of the accuracy with which the finite difference matrix equations are being solved. Despite this, it is not advisable to use this efficiency indicator alone because in some cases a small mass balance error may not indicate a high degree of model accuracy while, on the contrary, a large mass balance error is usually indicative of a low accuracy. If the mass balance error is equal to 0% it means that the mass is being perfectly conserved. Generally a mass balance error less than about 1% is an indication of good accuracy. If an error greater than 1% is obtained, it may indicate that the matrix solver is not solving the matrix equation with a good accuracy. To reduce mass balance error, the grid spacing and the time step have to be changed. There is just one case when a large mass balance error is not a sign of inaccuracy. It is the case in which there is a sudden change of pressure head in a boundary condition. This leads to an instantaneous change in the water stored in the mesh grid close to the boundary that was noted here in the step before so the model sees it as a discrepancy in the total mass balance. This situation happens in every first time step of a simulation because there is discrepancy between initial conditions and the first time step.

-Model Tolerance

The HYDRUS program numerically solves the Richards equation for saturated-unsaturated water flow. Due to the high non-linearity of Richards' equation, an iteration process is adopted to solve the global matrix equation for each time step. Using the Gaussian elimination solving method or the conjugate gradient method the system of algebraic equations derived from the global matrix is solved. Then the global matrix is inverted and its new coefficients are recalculated using the first solution then the process is started again for the next time step. The iteration stops when in every point of the saturated and unsaturated domain changes in the two main variables (water content and pressure head) are less than the value imposed as water content/pressure head tolerance. It is important to set a tolerance that is representative of the investigated variables and of the precision of the instrumentation. For example a reasonable tolerance for the water content is between 0,01 and 0,05 while for pressure head between 0,05 m to 0,1m. When setting the tolerance of a physical property it is extremely important to take into account the precision of the measuring devices in use. In our case GS3 and SM150T with a specific calibration have a precision of +/-0,01-0,02 and with a general calibration a precision of 0,03-0,04. T8 has an accuracy of +/-0,5 kPa (0,05 m) in the working range between -85/+100 kPa and MPS6 has an accuracy of -8 /+12 kPa (0,81-1,22m) in the range -9 to -100 kPa. The choice of a tolerance below 0,01 (-) for water content and 0,05 m for pressure head in our case has no significance.

2_7_2_3 MODEL SELECTION CRITERIA

Probabilistic statistical parameters could be used to quantify both the model performance on the training dataset and the model complexity. As a final goal they are able to guide the model selection ranking the analysed models (from the best to the worst). A large number of criteria have been introduced with this purpose, each of them lead to different theoretical properties and is suitable for a different type of database. Akaike and Bayesin Information Criterion are penalized-likelihood criteria extensively used to perform model comparisons. AIC works well with large dataset while BIC favors simpler model due to a tighter penalization for the number of parameters (with respect to AIC). In general, it is suggested to use both the criteria in model selection.

- AIC (Akaike Information Criterion)

To compare different models, the Akaike information criterion (AIC) (Akaike, 1974) could be used to compare the quality of the fit among the selected models while taking into account the number of parameters. For Gaussian process, AIC can be estimated from the residual sum of squares from the fitted model [equation 88]:

$$\mathrm{AIC} = \mathrm{n} \cdot \left[\ln(2 \cdot \pi) + \ln\left(\frac{\sum_1^{\mathrm{n}}(\mathrm{S_i} - \mathrm{M_i})^2}{\mathrm{n}-\mathrm{m}}\right) + 1\right] + \mathrm{m} \qquad \text{[equation 88]}$$

Where $''\mathrm{n}''$ is the number of recorded measurements and $''\mathrm{m}''$ is the number of estimated parameters. AIC could be seen as the sum of a constant and the relative distance between the unknown true likelihood function of the data and the fitted likelihood function of the model. A good model is the one that has minimum AIC among all the other models because a lower value of AIC means that the model is closer to the ''truth''. According to Burnham and Anderson (2004) the expression of AIC could be used only if the number of data points n divided by the number of parameters (m) is greater than 40. Models having AIC <2 have substantial support (evidence), between 4 and 7 have considerably less support, above 10 have no support at all.

It is possible to define a value of *AICmin* among all the indirect simulations performed with that same hydraulic model. A ranking between these simulations could be performed easily calculating the ratio *AIC/AICmin*. More this ratio is close to one, more the simulation is well evaluated.

-BIC (Bayesian Information Criterion)

$$\mathrm{BIC} = (\mathrm{AIC} - 2 \cdot \mathrm{m}) + \mathrm{m} \cdot \ln(\mathrm{n}) \qquad \text{[equation 89]}$$

Where $''\mathrm{n}''$ is the number of recorded measurements and $''\mathrm{m}''$ is the number of estimated parameters. Bayesian information criterion (BIC) is another criteria for model selection that measures the trade-off between model fit and complexity of the model. A lower BIC value indicates a better fit because the model is considered to be more likely the true model (Stone, 1979). BIC and AIC could assume also negative value in this case the best fit is the smallest negative value.

It is possible to define a value of *BICmin* among all the indirect simulations performed with that same hydraulic model. A ranking between these simulations could be performed easily calculating the ratio *BIC/BIC min*. More this ratio is close to one, more the simulation is well evaluated.

-HQIC (Hannan-Quinn Information Criterion)

$$\Phi = (\mathrm{AIC} - 2 \cdot \mathrm{m}) + \mathrm{c} \cdot \mathrm{m} \cdot \ln(\ln(\mathrm{n}))$$ [equation 90]

Where c is a constant greater or equal to 2 (Carrera & Neuman,1986).

This criterion (Hannan, 1980) measures the goodness of the fit of a statistical model. It is not based on a log-likelihood function but is related to the AIC Criterion. As seen for AIC and BIC, it assumes a penalty term for the number of parameters. A lower HQIC implies fewer explanatory variables of the model, a better fit or both.

It is possible to define a value of *HQICmin* among all the indirect simulations performed with that same hydraulic model. A ranking between these simulations could be performed easily calculating the ratio *HQIC/HCQICmin*. More this ratio is close to one, more the simulation is well evaluated.

2_7_2_4 STATISTICAL SIGNIFICANCE CRITERIA

Usually researchers are more interested in the difference between means than in the specific values of the means. If we assume that the null hypothesis Ho is: ''the differences between means are zero because they are purely due to random.'' Significance levels (α level) are used to test an hypothesis and they represent the probability of making the wrong decision when the null hypothesis is true. The confidence level (CL) is calculated as $(1 - \alpha \cdot 100)\%$, so if α=0,05, the corresponding confidence level is 95%. The CL 90%, 95%, 99% are the most commonly used. If a CL equal to 95% is adopted it means that we have the 95% probability that we choose the right hypothesis with a risk of error equal to (100-95)/100=5%.

- Student's t-test

Student's t-test is among the most commonly used statistical significance tests. It is able to compare the mean of two unrelated groups of samples. The t-test is able to evaluate if the means of two groups are statistically different from each other. First of all we have to formulate two kinds of hypotheses: the null hypothesis assumes that no differences exist between the two means, the alternative hypothesis assumes that the two means are statistically different. The final goal of this statistical test is to determine if the null hypothesis has to be rejected or not.

The two means and the corresponding standard deviations are calculated by using the following equations [equation 91, 92, 93, 94, 95]. n_X and n_Y are the number of measurements in dataset X and dataset Y, respectively:

$$\bar{X} = \frac{1}{\mathrm{n_X}} \cdot \sum_1^{\mathrm{n_X}} \mathrm{x_i}$$ [equation 91]

$$\bar{Y} = \frac{1}{\mathrm{n_Y}} \cdot \sum_1^{\mathrm{n_Y}} \mathrm{y_i}$$ [equation 92]

$$\sigma_X^2 = \frac{\sum_1^{\mathrm{n_X}} (\mathrm{x_i} - \bar{X})^2}{\mathrm{n_X} - 1}$$ [equation 93]

$$\sigma_Y^2 = \frac{\sum_1^{\mathrm{n_Y}} (\mathrm{x_i} - \bar{Y})^2}{\mathrm{n_Y} - 1}$$ [equation 94]

$$\sigma^2 = \frac{(n_X-1)\cdot\sigma_X^2+(n_Y-1)\cdot\sigma_Y^2}{n_X+n_Y-2}$$ [equation 95]

Finally, the statistic $\mathbf{t_{exp}}$ (experimental t value) is calculated:

$$t_{ex} = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{\sigma^2}{n_X}+\frac{\sigma^2}{n_Y}}}$$ [equation 96]

$$df = n_X + n_Y - 2$$ [equation 97]

$\mathbf{t_{ex}}$ value is compared with the **critical** (theoretical) $\mathbf{t_{th}}$ value corresponding to the given **degree of freedom df** and the chosen confidence level $(1-\alpha\cdot 100)\%$. The quantity α is the maximum acceptable risk of falsely rejecting the null-hypothesis. The smaller the value of α, the greater the strength of the test. Tables of critical t-values $\mathbf{t_{th}}$ can be found in any book of statistical analysis, as well as in many quantitative analysis textbooks. If $\mathbf{t_{ex}>t_{th}}$ then H_0 is rejected else H_0 is retained. However there is no need to use statistical tables containing critical values: it is sufficient to compute a numerical value *p*. The *p-value* is the probability of the error made by rejecting a null hypothesis when it is actually true.

For example, supposing that we have decided to work at Confidence level 95% ($\alpha = 0{,}05$)

A value of p = 0.085 means that **H_0 must be accepted**$\longrightarrow p\ value > \alpha$

A value of p = 0.021 means that **H_0 must be rejected** $\longrightarrow p\ value < \alpha$

A p-value of 5% or lower is often considered to be statistically significant.

Accordingly, if we decide to work at CL 90%, in both cases (P = 0.085, P = 0.021) H_0 is rejected, whereas if we decide to work at CL 99%, in both cases H_0 is accepted.

Graphically the *p-value* is the area underneath the curve (observed data on the x axis - probability density on the y axis) past the observed data point considered. p- value range between 0 and 1. If we are testing the possibility in just one direction, disregarding the possibility in the other direction, a one-tailed test is chosen. On the contrary if both the possibilities are investigated, the two-tailed test has to be chosen. In the one-tailed test, with a significance level of 0,05; all the significance is allotted in the single tail. For the two-tailed test, 0,025 in each tail.

- z-test

A z-test is a statistical test to determine whether two means are statistically different, as seen for student's t-tests. It is used when the variances of two populations are known and the sample size is large. While the t-test is based on Student's t-distribution, the z-test is based on the normal one. Both the distributions are symmetrical and with a bell-shape but the t-distribution is more ''narrow'' in the centre and ''wider'' in the tails. As seen for the t-test, two hypothesis are formulated (null and alternative hypothesis). If the statistic test is lower than the critical value, the null hypothesis is accepted or else it is rejected.

2_7_2_5 OTHER STATISTICAL INDICES

Uncertainty in measured data, boundary conditions and initial conditions have an adverse effect on the two statistical measures of 95% Confidence Limits and parameters correlation (Simunek et al, 2012). Both the statistical indices increase with uncertainties. Due to the high non linearity of the flow equation it is almost impossible to understand how the error will propagate in the modelling output or in the values of the optimized parameters. Both the statistical indices are obtained investigating the objective function in vicinity of its minimum, where the optimized parameter has been defined. Both express how well a parameter has been optimized or if it is possible to be optimized.

- Confidence Interval

The confidence interval is a type of estimate computed from the statistics of the observed data. It proposes a range of plausible values for an unknown parameter (for example, the mean). The interval has an associated confidence level that the true parameter is in the proposed range. Confidence level refers to the percentage of probability, or certainty, that the confidence interval would contain the true parameter. The confidence level of 95 % is usually selected because it is considered "good enough". With a 95% confidence interval the probability of observing a value outside of this area is less than 0.05.

The concept of the confidence interval is very important in statistics since it is used as a measure of uncertainty. Most frequently, we use confidence intervals to bind the mean. A confidence interval is a range around a measurement that conveys how precise the measurement is. The confidence interval tells you more than just the possible range around the estimate because it also gives information about how stable the estimate is. Wider confidence intervals in relation to the estimate itself indicate instability. Confidence intervals may also be used to determine whether two samples means are statistically different. The confidence interval is calculated based on the standard error of a measurement.

If we consider a Student's t-distribution the formula for the confidence interval can be expressed as:

$$\bar{X} = \frac{1}{\mathrm{n_X}} \cdot \sum_1^{\mathrm{n_X}} \mathrm{x_i} \qquad \text{[equation 98]}$$

$$\sigma_X^2 = \frac{\sum_1^{\mathrm{n_X}} (\mathrm{x_i} - \bar{X})^2}{\mathrm{n_X} - 1} \qquad \text{[equation 99]}$$

The t-distribution is described by df degrees of freedom:

$$\mathrm{df} = \mathrm{n_X} - 1 \qquad \text{[equation 100]}$$

$$\left[\bar{X} - t^* \cdot \frac{\sqrt{\sigma_X^2}}{\sqrt{n_X}} , \bar{X} + t^* \cdot \frac{\sqrt{\sigma_X^2}}{\sqrt{n_X}} \right] \qquad \text{[equazione 101]}$$

$$\frac{\sqrt{\sigma_X^2}}{\sqrt{n_X}} = SEC = standard\ error\ coefficient \qquad \text{[equazione 102]}$$

$$t^* = t^*(\mathrm{n_X}, \mathrm{df}, \alpha) \qquad \text{[equazione 103]}$$

Where t^* (α/2,df) is the upper critical value of the Student's-t distribution with df degrees of freedom to the desired significance level α (such that the probability of observing a value greater than t* is equal to $\boldsymbol{\alpha}$).

The confidence interval is obtained in Hydrus software by Pc-Progress in the output of the indirect problem from the Standard Error Coefficient (SEC) as follows:

Lower limit= parameter value – $Z_{0.95}$ *SEC [equation 104]

Upper limit= parameter value + $Z_{0.95}$ *SEC [equation 105]

Where ''parameter value'' is the value of the optimized parameter, $Z_{0,95}$ is the approximated value of 95 percentile point of the t-distribution.

If we consider a normal distribution the formula for the confidence interval can be expressed as:

$$\left[\overline{X} - z^* \cdot \frac{\sqrt{\sigma_X^2}}{\sqrt{n_X}}, \overline{X} + z^* \cdot \frac{\sqrt{\sigma_X^2}}{\sqrt{n_X}}\right] \quad \text{[equation 106]}$$

Where z^* (α/2,df) is the upper critical value of the normal distribution with df degrees of freedom to the desired significance level α.

-Mutual correlation between parameters

In statistic, the covariance provides a measure of the strength of the correlation between two or more variables. The sign of the covariance shows which type of linear relation links two variables. If a greater value of one variable corresponds to a greater value of the other, the covariance is positive, while if the variables show the opposite behaviour, the covariance is negative.

Covariance evaluates how much the considered variables change together. In other words, covariance is a measure of the variance between two variables without assessing the dependency between them.

If we consider two random variables:

$X = \{x_1, x_2, \ldots., x_n\} \qquad Y = \{y_1, y_2, \ldots., y_n\}$ [equation 107]

$\bar{X} = \frac{1}{n} \cdot \sum_1^n x_i$ (mean value) [equation 108]

$\bar{Y} = \frac{1}{n} \cdot \sum_1^n y_i$ (mean value) [equation 109]

$Cov\ (X, Y) = \frac{1}{n} \cdot \sum_1^n (x_i - \bar{X}) \cdot (y_i - \bar{Y})$ (covariance) [equation 110]

The variance is defined as the expectation of a variable from its mean.

$\sigma_X^2 = \frac{1}{n} \cdot \sum_1^n (x_i - \bar{X})^2 = COV(X, X)$ (variance) [equation 111]

Variance and covariance are often displayed together in a variance-covariance matrix.

The diagonal elements of the matrix contain the variances of the variables and the off-diagonal elements contain the covariances between all possible pairs of variables.

The variance-covariance matrix is symmetric because the covariance between X and Y is equal to the covariance between Y and X. This is the reason why the covariance for each pair of variables is displayed

twice in the matrix: the covariance between the i^{th} and j^{th} variables is displayed at positions (i, j) and (j, i) of the matrix.

Many statistical applications as the one implemented in Hydrus calculate the variance-covariance matrix for the estimators of parameters and often to compute the standard errors of estimators (see Standard Error Coefficient paragraph in this chapter §2_7_2_1 for more information). Hydrus software defines the covariance matrix C as follows:

$\mathrm{f}(\hat{\mathrm{x}}, \theta) = 0 \qquad \mathrm{f} \in \mathbb{R}^{\mathrm{m}} \hat{\mathrm{x}} \in \mathbb{R}^{\mathrm{m}} \theta \in \mathbb{R}^{\mathrm{n}} \quad \mathrm{m} > n$ [equation 112]

$\theta \in \mathbb{R}^{\mathrm{n}}$ parameter vector, $\hat{\mathrm{x}}$ measurement vector,

$\mathrm{r}(\theta) = \mathrm{f}(\hat{\mathrm{x}}, \theta)$(residuals) [equation 113]

$\mathrm{r_i} = \mathrm{f_i}(\hat{\mathrm{x}}, \theta) \quad \mathrm{i} = 1, \mathrm{m}$ [equation 114]

The LM algorithm minimizes the objective function **Φ** :

$\Phi(\boldsymbol{\theta}) = \sum_1^m r_i^2(\boldsymbol{\theta})$ [equation 115]

Starting with an initial guess $\boldsymbol{\theta^{(0)}}$ for the minimum, the method proceeds by iterations:

$\boldsymbol{\theta^{(k+1)}} = \boldsymbol{\theta^{(k)}} + \Delta\boldsymbol{\theta}$ [equation 116]

$\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right) \cdot \Delta\boldsymbol{\theta} + \boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{r} = \mathbf{0}$ [equation 117]

$(\boldsymbol{J_{res}})_{\boldsymbol{ij}} = \frac{\boldsymbol{\partial r_i(\theta^{(k)})}}{\boldsymbol{\partial \theta_j}}$ [equation 118]

$\Delta\boldsymbol{\theta} = -\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right)^{-1} \cdot \boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{r}$ [equation 119]

where $\boldsymbol{J_{res}}^{\boldsymbol{T}}$ denotes the matrix transpose.

The matrix $\left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right)$ on the left hand side is the approximate Hessian.

The residual standard deviation is defined as:

$\sigma_{res} = \sqrt{\frac{\sum r_i^2}{m-n}}$ [equation 120]

$df = degree\ of\ freedom = m - n;$

Under the stated assumptions, the parameter covariance matrix C is:

$C \approx {\sigma_{res}}^2 \cdot \left(\boldsymbol{J_{res}}^{\boldsymbol{T}} \cdot \boldsymbol{J_{res}}\right)^{-1}$ [equation 121]

To calculate the proper standard error for a given parameter, the following formula is applied:

$\sigma_{\theta_j} = \sqrt{C_{jj}} = \sigma_{res} \cdot \sqrt{H_{jj}}$ [equation 122]

where $\mathrm{H_{jj}}$ is the j-th diagonal element of the inverse of the Hessian matrix.

The Correlation Matrix R is defined using the elements of the covariance matrix. The elements in the Correlation Matrix could be computed as follows:

$$R_{IJ} = C_{ij}/(\sqrt{C_{ii}} \quad \sqrt{C_{JJ}} \qquad \text{[equation 123]}$$

The correlation matrix displays the correlation coefficients between the independent variables of a model. The correlation coefficients quantify the changes in the prediction due to the change of the input parameter i and they compare it to the change due to another parameter j. A value of +/-1 means a perfect correlation while a value of 0 means that the two parameters are perfectly uncorrelated. If there is an evident correlation between optimized parameters, it is strongly advised to optimize them singularly, in this way one is fixed and the other is free to change. If an optimization process of the two correlated variables is necessary, it is advisable to add new data for the inverse process. In some cases, levels of high correlation in the inverse analysis have led to problems of convergence and non-uniqueness of the solution.

- Simulation time

In a direct problem, the time necessary to model the whole simulation period can be an efficient parameter which helps to choose rationally the geometrical dimension of the mesh and of the model itself (Length*Width*Height), the constitutive mechanical and hydraulic models that could be adopted (in relation also to the objectives of the simulation) and the accepted tolerance in the calculus. In the inverse problem, the computational time is related to the number of iterations performed by the software which corresponds to the number of times the direct problem has been recalculated with a new set of initial parameters. A higher computational time means more iterations and a higher probability to have an estimated value of the optimized parameter significantly different from the initial one. A lower computational time means that the parameter slightly deviates from the initial value that is very close to ''the best'' value to be chosen with this set of initial parameters and the chosen input data. In the indirect problem, time convergence is not a relevant parameter to be considered in the calibration procedure.

In the following Table 1 a summary of some of the introduced metrics and their best value is presented:

Table 1: Summary of some metrics/indices that will be used in the following chapters, with indication of the mathematical formula and of the best fit value.

Metric	Mathematical formula	Description	Best value
MAE	$MAE = \frac{1}{n} \cdot \sum_{1}^{n} \lvert M_i - S_i \rvert$	*Mean Absolute Error*	0
MBE or ME	$MBE = \frac{1}{n} \cdot \sum_{1}^{n} (M_i - S_i)$	*Mean Bias Error*	0
MSE	$MSE = \frac{1}{n} \cdot \sum_{1}^{n} (M_i - S_i)^2$	*Mean Square Error*	0
RMSE	$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{1}^{n} (M_i - S_i)^2}$	*Root Mean Square Error*	0
CRM	$CRM = \frac{(\sum_1^n M_i - \sum_1^n S_i)}{\sum_1^n M_i}$	*Coefficient of Residual Mass*	0

SEC	$\mathrm{SEC}=\frac{6}{\sqrt{n}}$	*Standard error coefficient*	
NCV	$\mathrm{NCV}=\left(\frac{\sigma_X}{\bar{M}}*100\right)\%$	*Normalized Coefficient of variation*	0
R	$R=\frac{\sum_1^n(M_i-\bar{M})(S_i-\bar{S})}{\sqrt{\sum(M_i-\bar{M})^2}\cdot\sqrt{\sum(S_i-\bar{S})^2}}$	*Pearson's correlation coefficient*	1
R^2	$R^2=\left\{\frac{\sum_1^n(M_i-\bar{M})(S_i-\bar{S})}{\sqrt{\sum(M_i-\bar{M})^2}\cdot\sqrt{\sum(S_i-\bar{S})^2}}\right\}^2$	*Coefficient of determination*	1
r^2	$r^2=\frac{(\sum w_iM_iS_i-(\sum w_iM_i*\sum w_iS_i)/\sum w_i)^2}{(w_iM_i^2-\frac{(\sum w_iM_i)^2}{\sum w_i})\,(w_iS_i^2-\frac{(\sum w_iS_i)^2}{\sum w_i})}$	*Coefficient of determination considering the internal weights*	1
NSE	$NSE=1-\frac{\sum_1^n(M_i-S_i)^2}{\sum_1^n(M_i-\bar{M})^2}$	*Nash-Sutcliffe efficiency*	1
NSE log	$NSE_{log}=1-\frac{\sum_1^n\left(M_i^{log}-S_i^{log}\right)^2}{\sum_1^n\left(M_i^{log}-\overline{M^{log}}\right)^2}$ $M_i^{log}=log(M_i+c)$ $S_i^{log}=log(S_i+c)$	*log Nash-Sutcliffe efficiency*	1
IA	$IA=1-\frac{\sum_1^n(M_i-S_i)^2}{\sum_1^n(\lvert S_i-\bar{M}\rvert+\lvert M_i-\bar{M}\rvert)^2}$	*Index of Agreement*	1
NSE_j	$NSE_j=1-\frac{\sum_1^n\lvert M_i-S_i\rvert^j}{\sum_1^n\lvert M_i-\bar{M}\rvert^j}\quad j\in N$	*Modified efficiency*	0
IA_j	$IA_j=1-\frac{\sum_1^n\lvert M_i-S_i\rvert^j}{\sum_1^n(\lvert S_i-\bar{M}\rvert+\lvert M_i-\bar{M}\rvert\ )^j}\quad j\in N$	*Modified Index of Agreement*	1
NSE_{rel}	$NSE_{rel}=1-\frac{\sum_1^n\left(\frac{M_i-S_i}{M_i}\right)^2}{\sum_1^n\left(\frac{M_i-\bar{M}}{\bar{M}}\right)^2}$	*Relative efficiency*	1
IA_{rel}	$IA_{rel}=1-\frac{\sum_1^n\left(\frac{M_i-S_i}{M_i}\right)^2}{\sum_1^n\left(\frac{\lvert S_i-\bar{M}\rvert+\lvert M_i-\bar{M}\rvert}{\bar{M}}\ \right)^2}$	*Relative index of agreement*	1
$IA_{j=1}$	$IA_{j=1}=1.-\frac{\sum\lvert M_i-S_i\rvert}{\sum(\lvert S_i-\bar{M}\rvert+\lvert M_i-\bar{M}\rvert)}$	*Modified index of agreement with j=1*	1
IA'	$IA'=1.-\frac{\sum\lvert M_i-S_i\rvert}{2*\sum\lvert M_i-\bar{M}\rvert}$	*Modified index of agreement*	1
KGE	$KGE=1-\sqrt{(R-1)^2+\left(\frac{\sigma_S}{\sigma_M}-1\right)^2+\left(\frac{\bar{S}}{\bar{M}}-1\right)^2}$	*Kling-Gupta efficiency*	1
AIC	$\mathrm{AIC}=\mathrm{n}\cdot\left[\ln(2\cdot\pi)+\ln\left(\frac{\sum_1^{\mathrm{n}}(\mathrm{S_i}-\mathrm{M_i})^2}{\mathrm{n}-\mathrm{m}}\right)+1\right]+\mathrm{m}$	*Akaike Information Criterion*	
BIC	$\mathrm{BIC}=(\mathrm{AIC}-2\cdot\mathrm{m})+\mathrm{m}\cdot\ln(\mathrm{n})$	*Bayesian Information Criterion*	

HQIC	$\mathrm{HQIC} = (\mathrm{AIC} - 2 \cdot \mathrm{m}) + \mathrm{c} \cdot \mathrm{m} \cdot \ln(\ln(\mathrm{n}))$	*Hannan-Quinn Information Criterion*	

The model tolerance, the t-student test, the z-test, p-value, mass balance error, confidence intervals and correlation coefficients are not reported in the previous table because they could not be described by a single formula as the reported metrics.

3 THE CASE STUDY

Secchia River is the main right hand tributary of Po River after Panaro, whose catchment area covers 2292 km^2 and its length reaches 172 km, the 57% of its course is in mountain environment and it represents 3% of the entire Po basin. As all Appennine watercourses, Secchia River alternates summer lean with spring and autumn flash floods. In fact, due to the proximity with the mountains, the basin catchment has a rapid response to precipitations. The particularly violent autumn floods are partly controlled in the upstream stretch of the river (in Campogalliano, near Modena) by a complex system of reservoir dams involving a surface of about 1,000 hectares, with a water capacity of about 15 million cubic meters. In this regime just a small amount of water could infiltrate in riverbanks, and for this reason the stability could be guaranteed by the unsaturated conditions. A change of this regime could cause stability problems since a persistent high water level could result in a deep saturation front leading to a possible failure.

Secchia River embankments have a natural origin, but starting from XIX century anthropic actions changed the course of the river and its geometric sections in some parts. An overall look to River Secchia embankments shows earthly structures of considerable height (more than 7 m) with a crest 4-8 m wide, hillslopes with angles above 30° and a width of the base of less than 10 m.

The research activities that are going to be presented in the following chapters are part of the INFRASAFE project funded in the POR FESR 2014-2020 scheme. The main goal of the project is the design and realisation of an integrated system to characterise and to continuously monitor river embankment stability under transient seepage conditions. A full scale geotechnical monitoring system has been implemented in a section of the Secchia River. The in situ installation of instruments and the majority of the tests performed to characterise physically, mechanically and hydraulically the soil samples have been performed by a research group, in particular by doctor Gragnano (DICAM, Bologna University) and doctor Rocchi (Department of Civil Engineering, Denmark University). A number of publications on the topic could be found in literature, such as Rocchi et al (2018a), Rocchi et al (2018b), Gragnano et al (2018), Gragnano et al (2019).

No information are available on the study section about starting geometry, original course of the river and shape changes in time (Rocchi et al, 2018a). The chosen embankment section is 10 km downstream the location where during the flood event of January 19th, 2014 a disastrous collapse of a portion of the right bank of Secchia River occurred (D'Alpaos et al, 2014). 38 million of cubic meters of water come out of the breakthrough causing one casualty and huge economic damages. An accurate inspection by high-resolution aerial photographs post flooding showed a number of animal burrows that most likely triggered the event (Vacondio et al, 2015). The location of the monitored section was chosen in the same portion of the river

where in 2014 remediation works were carried out by AIPO (Caleffi and Cerutti, 2014). Moreover piezometers and some sensors to monitor water content were installed during these interventions. The new sensors were installed in an area upstream to avoid any influence of the past interventions (see Figure 2).

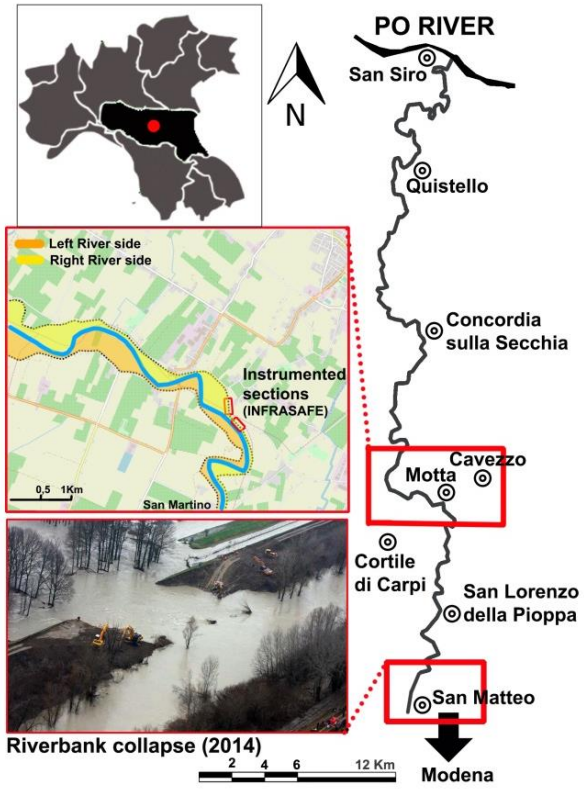


Figure 2: schematic representation of Secchia watercourse with indication of the location of the embankment collapse in 2014 and the new monitored section close to Cavezzo, Modena.

The monitored embankment is 11 m high from the ground level, the inner slope is 32° and the outer slope is 28°. The crest of the embankment is 4,6 m wide with a white road used for the maintenance and emergency vehicles. The berm is 5,5 m wide and positioned 5,2 m below the crest of the embankment. In the vicinity of the berm, superficial landslides of little relevance occurred in the past, some of them were fixed by the AIPO remediation works previously mentioned. The crest is 38 m from the centre of the river. A topographic survey has been used to trace the embankment section. The inner and outer slopes are covered by low herbaceous vegetation, kept mowed during the whole year (mowed twice a year). The geometry of the investigated section is presented in Figure 3 below.

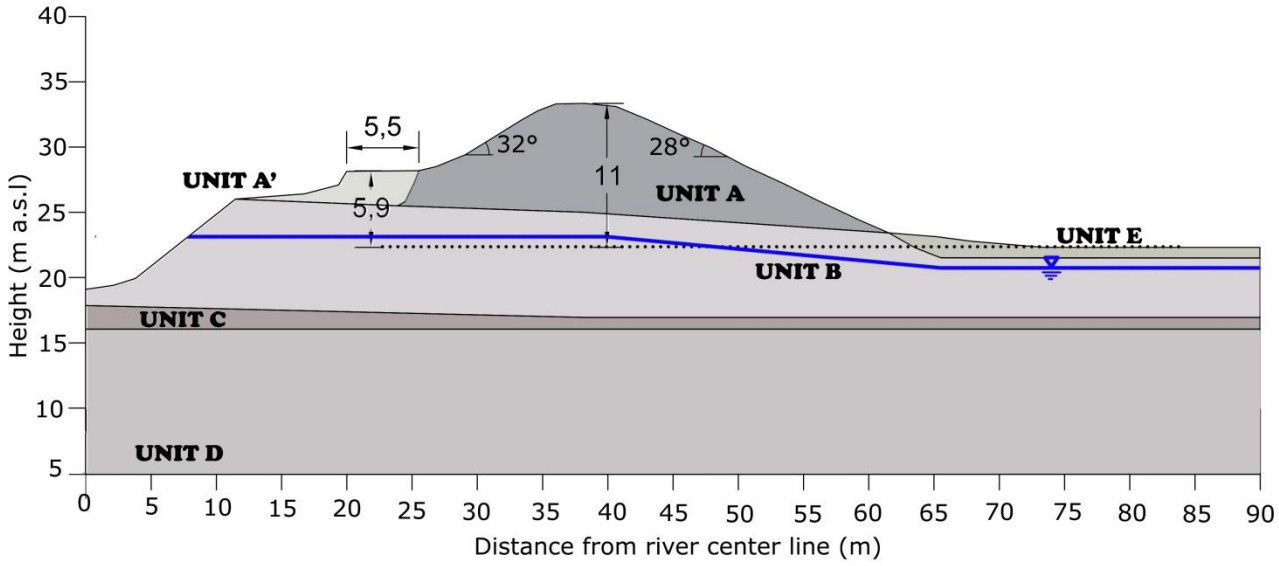


Figure 3: Schematic geometry of the investigated section of the instrumented embankment

The stratigraphy of the embankment has been investigated with a series of 4 CPTU tests, two in the berm and two in the crest reaching 15 m and 25 m respectively. q_t (corrected cone resistance) and u_2 (pore water pressure) data were interpreted using Robertson (2009) charts in terms of SBT (soil behaviour type). The elaborated stratigraphy shows an embankment constituted of fine grained material, a mix of silt and clay in different proportions, becoming more fine moving from the river to the embankment. The 4 CPTU tests show also a modest change in the depth between different layers in different verticals. Unit A is an inhomogeneous alternation of silt and sandy silt, unit A' consists of coarser fluvial sediments deposited during floods, unit B is made of a finer material compared to the above layer (unit A); unit C is a slightly coarser layer and it corresponds to the shallow aquifer affected by the changing hydrometric level of the river and rainfall contribution, with greater permeability compared to other layers, unit D, below unit C, is almost an uniform clayey layer. From the two continuous corings performed in the berm, the use of soil mixed with lime was found at certain depths, result probably of the previously mentioned remediation works, this hypothesis has been confirmed also by laboratory tests (carbonate content measurements) on the extracted cores. The 4 CPTU tests were performed during summer (June) and the embankment was highly in unsaturated conditions at the time, with a water table at the depth of 9 m from the crest. For this reason Q_t (the normalised cone resistance), F_r (normalised friction) and B_q (pore pressure parameter) were calculated using the stress state under the hypothesis that the suction profile was linear and equal to the one experienced in the same embankment the year later the sensors installation.

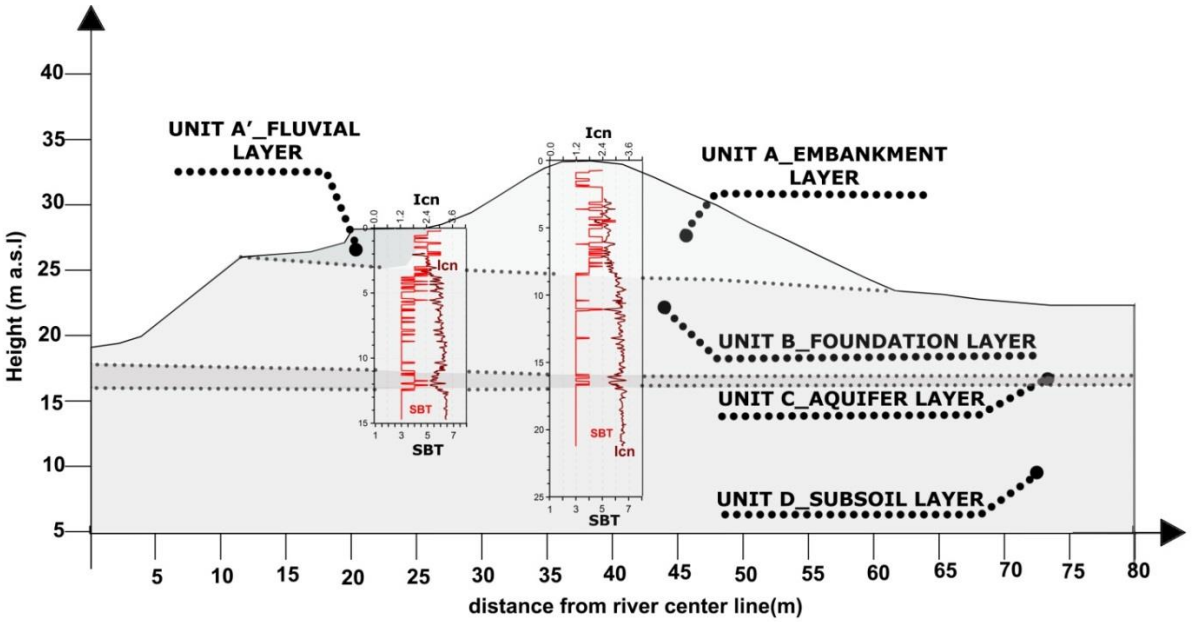


Figure 4 The geometry of the investigated section with indications of the different layers as obtained by the interpretation of the CPTU tests according to Robertson, 2009.

Characterization tests have been performed by doctor Gragnano and doctor Rocchi on the soil retrieved from the holes in which instrumentations were installed and on undisturbed samples. 9 grain size distributions, 7 specific gravity tests, 7 Atterberg limits were performed at different depths inside Unit A (embankment). 4 grain-size distribution, 4 specific gravity tests, 4 Atterberg limits, 4 organic contents for Unit A' (fluvial deposit) and 2 grain-size distribution, 2 specific gravity tests, 2 Atterberg limits, 2 organic contents for Unit B (foundation). In 5 undisturbed samples, the natural water content, the void ratio and the natural weight of the soil were obtained (Gragnano et al, 2018; Rocchi et al, 2018a). In Table 2, a summary of the results obtained from the characterization tests performed for the different soil units is presented:

Table 2: A summary of the results of the characterisation tests performed on the investigated layers.

	w_L (%)	w_P (%)	clay (%)	silt (%)	sand (%)	gravel (%)	w nat (%)	Gs (-)	organic content (%)	$CaCO_3$
embankment	32,9	19,3	13,7	48,0	38,3	0	18,4	2,646		
fluvial	44,2	19,7	20,2	57,6	14,5	7,8		2,598	1,72	5,72
foundation	40,7	20,4	16,5	58,9	23,2	1,3		2,606	1,53	5,36

3_1 THE INSTALLED INSTRUMENTATIONS

All the installed sensors are commercially sold by notorious company and price, accuracy, ease of installation and ease of calibration were the main characteristics to guide the choice. These sensors have been developed in the agricultural field in order to monitor, control and increase the level of efficiency in agricultural productivity on large areas (irrigation schedule, water availability for plants etc). The adopted sensors were applied with due care to engineering purposes, overcoming issues related to installation, maintenance and long-time data acquisition.

Between 2016 and 2017 the study section was instrumented with 7 GS3 by Meter Group (5 in the crest and 2 in the berm) and 4 SM150-T by Delta-T Device (3 in the crest and 1 in the berm) for volumetric water

content measurements, 2 Casagrande Piezometers, 9 MPS6 tensiometers by Decagon Device (6 in the crest and 3 in the berm) and 3 T8 tensiometers by UMS (2 in the crest and 1 in the berm) for pore water pressure measurements (in the positive and negative range). 20 sensors in total have been installed in the embankment body. T8 installation depth ranges from 5-7,9 m in the crest and 4,9m in the berm (just one sensor); GS3 installation depth ranges from 1,4-7,1 m in the crest and 0,7-2,2 m in the berm; SM150T has a range of 1,2-7,1 m in the crest (no sensor in the berm) and MPS6 has a range between 3,1-7,0 m in the crest and 0,9-2,7 m in the berm. The sensors have been installed in three times. In 2016 the Casagrande piezometers in the crest at a depth of 10 m and 17 m were installed by Hypo. Later in 2016, TC2, MPC2, MPC3, SPC3, SPB1 boreholes and the associated sensors installation were performed, in 2017 the remaining boreholes. A schematic drawing of the instrumented section is presented in Figure 5 below.

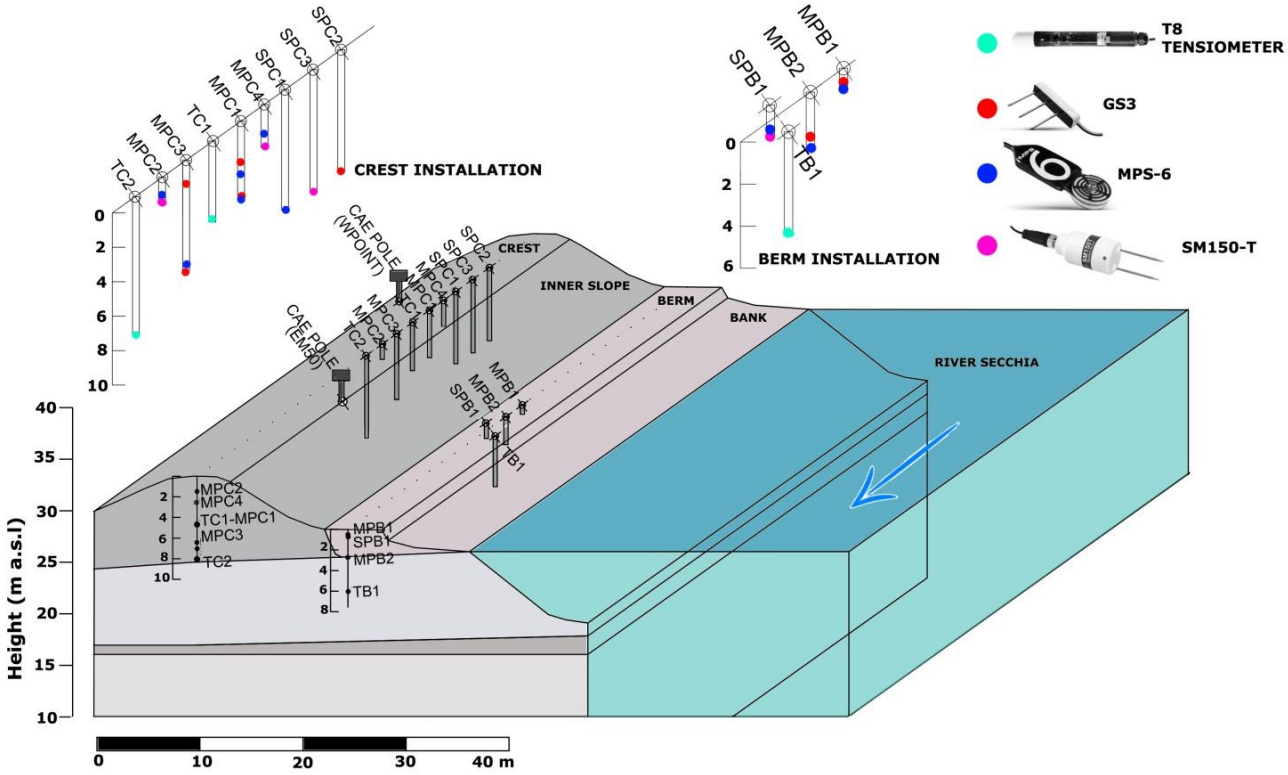


Figure 5 Schematic drawing of the instrumented section with indications of the boreholes and the installed sensors.

T8 measures directly matric water potential against atmospheric pressure in the working range between -85/+100 kPa, with a nominal measurement accuracy of +/-0.5 kPa (UMS, 2011). The atmospheric pressure is taken at the ground level and conducted through a diaphragm positioned on the cable to the pressure sensor. If the water potential goes below -85 kPa, the ceramic cup gets dry and it could be refilled trough two capillary tubes without being removed from the soil when the soil is sufficiently moist again. T8 tensiometer does not require a site specific calibration.

MPS-6 measures the dielectric permittivity of a porous ceramic disc in equilibrium with the surrounding soil volume (Decagon Device, 2016a). The soil dielectric permittivity is a parameter directly correlated to soil water content and indirectly to soil water potential by means of the proper SWRC. Every ceramic stone has the same pore size distribution, with pores ranging in a wide interval, in this way the ceramic stone drains in a wide water potential range. The calibration curve of the porous ceramic disk (moisture characteristic curve of the ceramic) is provided by the manufacturer. MPS6 has a 6 points calibration curve: one point at 0

kPa (saturated state), one point at -100 MPa (air-dry state) and 4 points between -9 and -100 kPa. The accuracy in the range -9 to -100 kPa is +/- 10% of the reading+2 kPa. Due to the fact that MPS-6 sensors were developed for agricultural purposes and the water content and suction range under investigation is quite different from the one of interest in river embankment, a laboratory calibration of MPS6 sensor's response was performed. To do so, MPS6 readings and tensiometer readings were acquired using the following procedure. The in situ soil is prepared at a target void ratio with the two sensors installed at the same depth, then a gradient is established inside the soil sample from the bottom to the top, then the surface exposed to atmosphere is covered and when equilibrium is reached the two lectures are acquired (raw signal from MPS6 and suction from the tensiometer). The procedure is repeated 9 times drying every time a little more the soil sample up to a suction of 60 kPa. To obtain the calibration curve for suction values above 1 MPa, tensiometers reading are substituted with Dewpoint Potentiameter (WP4) readings. Comparing the obtained calibration curve with the one provided by the manufacturer is it possible to notice a reduction of the difference between the two calibration curves moving from low to high suction values.

GS3 (Decagon Device, 2016b) and SM150 T (Delta-T Device, 2016) employ an oscillator to generate an alternating electrical current between two electrodes and measuring the resistance given by the surrounding soil. The sensor determines the volumetric water content (vwc) measuring the dielectric constant (εa) of the medium applying the frequency domain technology. This is based on the fact that soil resistance is extremely sensitive to water content variations (Bittelli, 2011). A magnetic field of 70 MHz is created around the 8 mm long metal rods of the GS3 sensor while a magnetic field of 100 MHz for the SM150T sensor. The two electromagnetic frequencies are chosen in order to minimize textural and salinity effects. The working range of both instrumentations is 0,0-1,0 m^3/m^3, with an accuracy of +/- 0,03 m^3/m^3 for a generic calibration (accuracy can rise up to +/- 0,01-0,02 m^3/m^3 for a site specific calibration). Sensor's producer supplies a generic calibration equation that could work in all types of soils but a specific calibration is suggested to increase the accuracy of the data. For this reason, the calibration equation for the soil of interest was obtained in laboratory. The in situ soil was prepared at a given void ratio and water content (of interest for the future installation in the embankment), and a GS3 sensor was installed in it. The soil sample is wet from the top with a certain amount of water, then after an equilibration period, the water content is recorded. The procedure has been repeated multiple times. Due to the higher monetary cost of SM150T (with respect to GS3), only three sensors have been installed.

Sensors were installed in boreholes created ad hoc, in a single-point installation or in a multiple-point installation. Multiple point installations have the advantages to reduce excavation costs and the number of holes (to not undermine embankment stability with preferential routes to water infiltration). Moreover the combined presence of a sensor for the water content measurement very close to a sensor for water potential measurement allow to obtain a SWRC in situ. As pointed out by Bordoni et al (2017), the determination of SWRC in laboratory shows intrinsic limitations such as availability of few experimental points, shorter investigated range of water content and suction, environmental conditions often not representative of the natural in situ conditions, disturbance in the sampled soil. It is therefore also relevant to remember that one of the big problems related to the use of multiple-point installation is the possible leak of water along sensors' cables from one instrument to the other (Rocchi et al, 2018a).

In the investigated embankment, a single-point installation has been used only for SM150T and T8 because these sensors were installed deeper (compared to GS3 and MPS6) at the base of the borehole in the undisturbed soil because disturbance could cause significant error in the readings. To ensure no

disturbance, a borehole of 5 cm of diameter was created and the last 20 cm were excavated using a hand auger with a smaller diameter (Rocchi et al, 2018b; Rocchi et al, 2018a).

12 boreholes were realized in total and every hole has been named according to these abbreviations: SP stands for ''single point coring'', ''T'' for ''tensiometer''and ''MP'' stands for ''multiple point coring''.

Due to the fragility of MPS6 sensor's porous stones, the in situ installation was performed creating around the probe a ''soil cake'' using the soil from the embankment. The cake than is lowered in the borehole. To guarantee a good contact between GS3 needles and the undisturbed soil around the hole, a prototype of instrument (Q-bit) has been used to push the sensors in the soil converting vertical force to a horizontal force. More details on the two procedures could be found in Rocchi et al (2018a).

3_2 LABORATORY-MEASURED SWRC OF THE INVESTIGATED SOILS

In time many experimental methodologies have been developed in order to define SWRC parameters.

Direct approaches are based on the simultaneous measurement of water content and pressure head information that are subsequently fitted using a particular function. Examples of direct approaches are sand box, hanging column, pressure plates, pressure cells, evaporation tests among others (Bittelli and Flury, 2009). Data are fitted using specific analytical functions as the ones proposed by Brooks and Corey (1964), van Genuchten (1980), Fredlund and Xing (1994) Kosugi (1996) among others. In some cases, linear, cubic or other interpolation techniques are used to analyse laboratory data instead of analytical functions (Bitterlich et al, 2004). A big effort has been made in the last years to elaborate advanced hydraulic models that need the estimation of an ever-growing number of parameters. To a growing complexity of the models corresponds a growing complexity of the laboratory tests for the initial estimate of the hydraulic parameters. Indirect approaches are based on Pedotransfer functions (PTF) that use regression analysis or neural network approach to correlate input parameters to obtain in output the unknown hydraulic parameters using data from soil database (Caruso and Jommi, 2005). This method is adopted when large areas need to be investigated and a comprehensive direct laboratory characterisation is not possible. Usually the input parameters are easily, routinely and cheaply measured properties such as grain size distribution, natural water content, Gs, organic content etc. Many databases around the word were established for pedotransfer applications such as UNSODA, HYPRES, WISE, NRCS among other.

One of the simpler laboratory tests that could be performed is the evaporation method, introduced by Gardner and Miklich (1962) and then modified by Wind (1968). A popular version of the evaporation method is performed using HYPROP by Meter Group. The system allows the measurement of pressure head at two depths in a 5 cm high soil sample while water evaporates from its surface and water loss is determined by weighing the system. Several studies have tested the evaporation approach using this commercial product (e.g. Schelle et al., 2010; Zhuang et al., 2017 among others).The mini-tensiometers allow to cover a relatively wide range of pressure heads (down to -240 kPa). The soil sample has a volume of 250 cm^3 and the tensiometers are positioned at 1,25 cm and 3,75 cm from the base of the cylinder containing the sample. One of the major advantages of Hyprop is the complete automatization once the program is installed and no preselected hydraulic function is set to elaborate the row data (Bezerra Coelho et al, 2018). Using measured pressure heads, water contents and evaporation fluxes it is possible subsequently to obtain the SWRC and the hydraulic conductivity function K(h). It is not recommended to measure independently Ks, the saturated permeability, because close to saturation, the mini-tensiometers

are not able to register very small changes in water potential. This is particularly relevant for all type of soils with the exception for relative fine and intermediate soils (Bezerra Coelho et al, 2018). The strength of the evaporation method stays in the closeness of the procedure to determine hydraulic parameters to the natural phenomenon that occurs in real soils (Romano and Santini, 1999).

The evaporation phenomenon evolves in two stages during the test. Initially the evaporative flux from the upper surface is constant over time and the decrement of the soil weight is linear in time. This is due to the fact that the evaporation rate is independent from the hydraulic properties of the soil but, at the initial stage, it depends only on the external environmental conditions. The duration of this initial stage depends on the fact that the increase in pressure head's gradient compensates the decrease in the hydraulic conductivity K(h) (Romano and Santini, 1999). In time the water content near saturation decreases drastically together with K(h) and the water could not be delivered to the external surface at the same initial rate. The second stage sees the abandon of the initial linear curve time-water loss and data arrange on a concave-shaped curve. The experiment stops when the tensiometer in the lower position cavitates, breaking the hydraulic connection to the pressure transducer (Romano and Santini, 1999). Evaporation test could be performed with a continuous drying or by stages, in which a drying stage is followed by an equilibration stage. In a continuous stage it is possible to slow down the evaporation, wrapping the sample or placing the sample in a controlled humidity chamber. Cunningham (2000) and Boso et al (2003) showed that little differences could be found in the obtained SWRC using stage drying or a continuous drying. This leads to the conclusion that the obtained SWRC has a low dependence on the evaporation rate (Lourenco et al, 2007). It should also be said that the most accurate results are obtained with the stage drying procedure because the lack of equalisation could cause inaccuracies that depend on size of the soil sample, shape of the sample, dimension of the exposed surface, position of the tensiometer. If the tensiometer is in contact with the external surface it is more likely to obtain an incorrect water retention curve (Lourenco et al, 2007).

It is relevant to report that the parameter estimation technique by means of inverse analysis has been successfully applied to the laboratory evaporation technique since long time. In fact hydraulic parameters estimation could be obtained fitting the data suction-volumetric water content to a certain function (a soil hydraulic model), the so-called Wind's method, or obtained from numerical inversion (Simunek et al, 1998; Simunek et al, 1999). Ciollaro and Romano (1995) applied the inverse procedure to determine soil hydraulic parameters of a number of soils using uniquely a dataset from an evaporation experiment. A good agreement between parameters estimation by means of the Wind's method and the inverse technique has been obtained also by Simunek et al (1998). Practical information on the correct laboratory procedure, the minimum number of information required for a reliable and accurate inverse estimation of the parameters could be found in Hopmans et al (2002).

13 evaporation tests were carried out to investigate hydraulic properties of the embankment of interest (unit A), among which 3 tests were performed using intact soil samples and the rest using reconstituted samples at the void ratio determined from the in situ state at the considered depth ranging from 2,8 m to 6,9 m from the crest. 3 evaporation tests were carried out on intact soil from the fluvial layer (unit A') at a ranging depth from 2,15 to 2,27 m, one test on an intact sample from the foundation (unit B) and one test on a reconstituted sample using the soil from the agricultural field by the outer side of the embankment. Reconstituted sample started from a condition close to saturation then a slow evaporation rate has been imposed to guarantee a linear distribution of suction along the sample height (Gragnano et al, 2018).

To trace the dry end of the SWRC also beyond the cavitation limits of the micro-tensiometers, the evaporation tests were coupled to vapour pressure methods. In this method suction is determined equilibrating the sample in an enclosed chamber and subsequently measure the vapour pressure of the chamber and the sample saturation vapour pressure. Total suction is then computed using Kelvin equation. Dew point measurement is a ''primary measurement'' because it uses a fundamental thermodynamic relationship and a temperature measurement to obtain total suction (Campbell et al, 2007). The Dew Point Potentiameter (model WP4 by Decagon Device) has been used for the laboratory tests. It determines total suction measuring the dew point temperature of the chamber in equilibrium with the investigated sample. The temperature of a mirror, whose reflectance is monitored by an optical sensor, is lowered to the point in which dew occurs. The sample temperature is monitored by infrared thermometry while mirror temperature is monitored by a thermocouple. Dew point method is the most efficient method to measure suction in the range 1-400 MPa with an accuracy of 1% while for the wet end it is not a substitute of tensiometer measurement and axis translation (Campbell et al, 2007).

In the following graphs (Figure 6), the performed evaporation tests are presented in terms of volumetric water content (θ %)- suction (kPa). The points are the laboratory data and the black lines are the best fitting using van Genuchten (1980) model. In image A all the SWRC of unit A (embankment layer) are plotted together, in image C the ones of unit A' (fluvial layer), in image D the retention curves of unit B (foundation) and in image B of the agricultural layer. The embankment is the layer which is better characterized with a greater number of performed laboratory evaporation tests because an accurate determination of its hydraulic parameters is of primary importance for reliable pwp spatial distributions in time and future stability analyses.

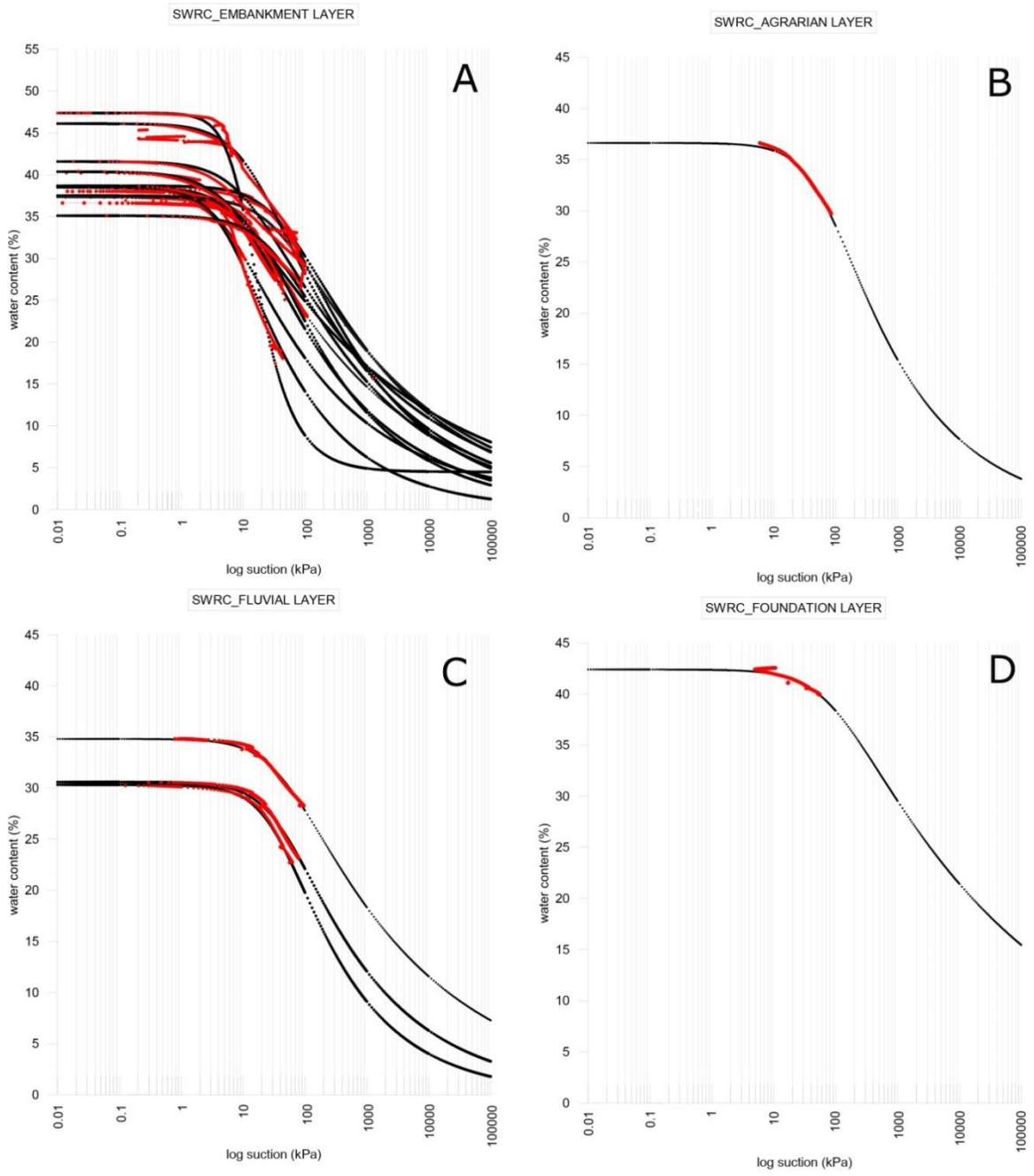


Figure 6: SWRC in terms of volumetric water content-suction of the investigated layers (image A the embankment layer, image B the agricultural layer, in image C the fluvial layer and in image D the foundation layer).

In Table 3 below, the hydraulic parameters obtained by laboratory evaporation tests have been reported divided for layer. Some statistical indices such as the mean, the standard deviation, the variance, the confidence interval have been computed for each layer. In red the statistical indices that show a higher dispersion of the optimized values (standard deviation and confidence interval of the same order of magnitude of the parameter) which indicate a greater uncertainty in the estimate.

Table 3: Hydraulic parameters obtained fitting the laboratory experimental data of the investigated soils and some statistical parameters (mean, standard deviation, variance and confidence interval).

EMBANKMENT	θR (-)	θsat (-)	αVG (m-1)	nVG (-)
TC1_2.8m (I)	0	0,375	0,903	1,207
TC1_2.8m (II)	0	0,385	0,827	1,265
TC1_2.8m (III)	0,011	0,385	1,830	1,268
SPC1_3.9m (I)	0	0,387	0,210	1,269
SPC1_3.9m (II)	0	0,387	0,232	1,223
SPC1_3.9m (III)	0	0,461	0,734	1,237
TC2_4.3m (I)	0,00008	0,375	0,491	1,300
SPC1_4.3m (II)	0,045	0,474	0,912	2,023
SPC1_4.3m (III)	0,0008	0,386	1,598	1,360
SPC1_4.8m (I)	0	0,374	0,268	1,204
SPC1_4.8m (II)	0	0,404	1,819	1,164
SPC1_4.8m (III)	0	0,351	0,321	1,238
SPC1_6.9m	0	0,416	0,909	1,196
mean	0,004	0,397	0,850	1,304
stand. deviation	0,013	0,035	0,579	0,222
variance	1,6E-04	0,001	0,335	0,049
n° values	13	13	13	13
confidence interval	0,008	0,021	0,350	0,134
average+conf.interval	0,012	0,418	1,200	1,438
average-conf.interval	0,000	0,376	0,500	1,170
FLUVIAL	θr (-)	θsat (-)	αVG (m-1)	nVG (-)
TB1_2,15 m	4,2E-06	0,348	0,233	1,201
TB1_2,27 m	8,0E-04	0,306	0,247	1,288
TB1_2,27 m (II)	8,0E-04	0,303	0,271	1,363
mean	0,001	0,319	0,251	1,284
stand. deviation	4,6E-04	0,025	0,019	0,081
variance	2,1E-07	6,3E-04	3,7E-04	6,6E-03
n° dati	3	3	3	3
intervallo confidenza	0,001	0,063	0,048	0,201
media+int conf	0,002	0,382	0,298	1,485
media - int conf	0	0,257	0,203	1,083
FOUNDATION	θr (-)	θsat (-)	αVG (m-1)	nVG (-)
TB1_4,8 m	0	0,424	0,118	1,142

3_3 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

Chapter §3 focuses on the presentation of the instrumented bank section along Secchia River, close to Cavezzo, Modena. The installed sensors and their functioning have been extensively discussed together with the accuracy of the instruments and the calibration procedures adopted (§3.1). Moreover, the laboratory tests performed in order to investigate the SWRC of the different layers have been described and the results shown (Figure 6, Table 3 of §3.2).

25 sensors monitoring soil water content or pwp have been installed in the investigated riverbank section (n° 7 GS3 by METER Group; n°4 SM150T by Delta-T Device; n°2 Casagrande piezometers;n°9 MPS6 by Decagon Device; n°3 T8 by Decagon Device) in the period between 2016-2017 (see Figure 5). These instrumentations have been designed and distributed for agricultural purposes in order to increase the efficiency of the productivity for very extensive crops. Their functioning has been converted to geotechnical engineering purposes. In some cases sensors have been installed individually in boreholes, in other cases multiple installations have been performed in the same hole (a sensor monitoring pwp and one monitoring water content) in order to track the in situ SWRC at a certain depth. T8 sensor measures directly the matric

water potential in the working range between +85/-100 kPa with a nominal measurement accuracy of +/- 0.5 kPa. MPS6 measures the dielectric permittivity of a porous ceramic disc in equilibrium with the moisture of the surrounding soil and the dielectric permittivity is directly related to soil water content. The accuracy of MPS6 sensor is +/- 10% of the reading +2 kPa in the range -9 to -100 kPa. In GS3 and SM150T sensors, two electrodes generate an alternating electrical current and the FDT is applied to measure vwc of the surrounding soil. The hydraulic parameters of the SWRC have been obtained by means of evaporation tests using Hyprop by METER Group (§3.2). The pwp in time is measured at two different heights inside the soil sample by micro-tensiometers (working range +2 to -240 kPa) while the water loss is determined by weighing the whole system during the test. Eighteen evaporation tests have been performed some of them on undisturbed samples, others on reconstructed samples (at the in situ void ratio). Tests have been performed starting from a condition close to saturation then samples have been gradually dried. The dry end of the drying branch of the SWRC (up to 1MPa) has been determined by means of the dewpoint method (WP4 by Decagon Device). Globally n°13 tests have been performed on the embankment soil, n°3 on the fluvial soil, n°1 on the foundation soil and n°1 on the agrarian soil (Figure 6). Statistical indices have been used to study the dispersion of the hydraulic parameters for each investigated layer (Table 3).

Chapter §4 presents the preliminary tests on the metrics/indices that could be potentially used to quantify the performance of direct and indirect simulations (inverse analysis) of the investigated problem. Metrics/indices have been tested using datasets similar to the ones with which we are dealing in the present case study (dataset of PH, dataset of wc and mixed dataset PH+wc) with an increasing dispersion of the data and an increasing number of outliers. The performance of observed-simulated datasets has been evaluated and the pros/cons of each index highlighted.

4 PRELIMINARY TESTS ON THE METRICS/INDICES PERFORMANCE

The metrics performance has been tested prior to their application to the case study. It is becoming increasingly important to test indicators in order to individuate the ones that are logical, consistent and appropriate for the application (Ali et al, 2014). The main characteristics that a good indicator should have are the following:

1) To have direct physical meaning and interpretation.

2) To catch the accuracy and weakness of the simulation capability, to allow, as final goal, an estimation of the acceptability/goodness of the model.

3) The trend of the metric has to be consistent with the logical directions without ambiguous performances. An index has to be always consistent in its result with different initial dataset in order to allow the comparison of different models or simulations.

4) Indices that are not dependent on the measure unit are more powerful because they are applicable to any field of observation and range of values. For our case study this characteristic is extremely useful due to the fact that we deal with pressure head and water content data that have usually different order of magnitude.

5) Bounded indices (with upper and lower values) have to be preferred. Many indices are bounded on the upper end (usually 1) but lack of a finite lower bound which makes very difficult the assessment and comparison of models performance.

To test the behaviour of indices/metrics, simulated random data have been used while observed data are real and taken from the in situ measurements. In particular for the observation data of typology ''water content'' SPC2_7,1m has been used (observed data in the period 147450-319950 min) while for the typology ''pressure head'' T8-TB1-4,9m has been used (observed data in the period 62400-135000 min). A third dataset from the combination of the pressure head and water content datasets has been used. 27 data are in the water content dataset, 30 data in the pressure head dataset and 57 in the pressure head+water content dataset. Random data have been generated by means of Matlab by Mathworks for each typology of observation dataset according to these criteria:

RANDOM 1_WATER CONTENT DATASET: random values in the interval 0,24-0,36%. The chosen interval is a possible range of variance of the water content of the soil.

RANDOM 2_WATER CONTENT DATASET: random values in the interval 0,22-0,40% (larger interval with respect to random_1).

RANDOM 3_WATER CONTENT DATASET: random values in the interval 0,19-0,60%(larger interval with respect to random_2).

RANDOM 4_WATER CONTENT DATASET: the same values of Random_2 but one of the values has been substituted by an outlier.

RANDOM 5_WATER CONTENT DATASET: the same values of Random_3 but four values have been substituted by outliers.

RANDOM 6_WATER CONTENT DATASET: all the random values are underestimated with respect to the observation data (all the points are below the 1:1 line in the graph of Figure 8)

RANDOM 7_WATER CONTENT DATASET: all the random values are overestimated with respect to the observation data (all the points are above the 1:1 line in the graph of Figure 8)

RANDOM 1_PRESSURE HEAD DATASET: random values in the interval 0,8-1,2m. The chosen interval is a possible range of variance of the soil pressure head.

RANDOM 2_ PRESSURE HEAD DATASET: random values in the interval 0,4-1,4m (larger interval with respect to random_1).

RANDOM 3_ PRESSURE HEAD DATASET: random values in the interval 0,2-2,0m (larger interval with respect to random_2).

RANDOM 4_ PRESSURE HEAD DATASET: the same values of Random_2 but one of the values has been substituted by an outlier.

RANDOM 5_ PRESSURE HEAD DATASET: the same values of Random_3 but three values have been substituted by outliers.

RANDOM 6_ PRESSURE HEAD DATASET: all the random values are underestimated with respect to the observation data (all the points are below the 1:1 line in the graph of Figure 8)

RANDOM 7_ PRESSURE HEAD DATASET: all the random values are overestimated with respect to the observation data (all the points are above the 1:1 line in the graph of Figure 8)

RANDOM 1_PRESSURE HEAD+WC DATASET: combination of Random_1_pressure head+ Random_1_water content datasets.

RANDOM 2_ PRESSURE HEAD+WC DATASET: combination of Random_2_pressure head+ Random_2_water content datasets.

RANDOM 3_ PRESSURE HEAD+WC DATASET: combination of Random_3_pressure head+ Random_3_water content datasets.

In Figure 7 (left) Random_1, Random_2, Random_3 pressure head datasets could be observed with their dispersion with respect to the black line (the real observation points of T8-TB1-4,9m). In Figure 7 (right) Random_1, Random_3, Random_3 water content datasets could be observed superimposed on the curve of the real observation points (SPC2_7,1m). In Figure 8 the random datasets of pressure head and water content are represented against the observed dataset, line 1:1 represents the best fit between simulated and observed data. Table 4 reports the indices/metrics values obtained for each random distribution presented. The values that are not consistent with the expectations are underlined in red colour.

- The difference-based statistical indicators as MAE, MSE, RMSE show consistent and logical values in all the different dataset typology (water content, pressure head, water content+pressure head). The error values increase as values become more scattered from Random_1 to Random_3, moving from 0 (best fit) to larger positive values. Moreover the dataset with outliers show an increase in the error compared to the simulation without outliers which is logical. It is important to report that the index based on absolute values of differences (as MAE) are in general preferable over the index based on squared differences (as RMSE) because squaring the errors prior to sum them over-weights the influence of bigger errors and under-weights the influence of smaller error (Willmott et al, 1985). This is a general statement that could be extended to all the indices.

- The index MBE shows to be inconsistent for the pressure head dataset, incrementing for Random_2_PH (going to positive value) and decrementing (going to negative value) for Random_3_PH. This is due to the fact that MBE does not have a quadratic form or an absolute value so the errors of opposite sign cancel out. The index MBE is able to show the general direction of the error (underestimation or overestimation of the observation dataset) and for this reason, it could give precious information if analysed together with the other error indices.

- The coefficient of determination R^2 shows an illogical behaviour for all the three typologies of dataset (pressure head, water content, pressure head+water content), increasing in Random_2 with respect to Random_1, and decreasing again in Random_3. The expected behaviour is a decrease proportional to the dispersion of the dataset (a proportional departure from the best fit, closer to 1). The correlation and correlation based indices such as R^2 are extremely sensitive to extreme outliers and insensitive to additive and proportional differences between observed and

simulated dataset (Willmott 1981) for this reason their use has always to be very cautious and always coupled with the use of other indices.

- The efficiency metrics such as NSE, NSErel and NSEj show a consistent and logic behaviour in all the three typologies of dataset. Despite the sensitivity of these efficiency criteria to outliers, the indicators give the expected response also for Random_4 and Random_5. NSE, NSErel and NSEj have negative values that mean that the average value of the observed dataset is a better predictor than the simulated dataset. The efficiency of the water content dataset is lower than the one of the pressure head dataset, this is due to the fact that NSE overestimates the efficiency of the dataset with bigger numerical values (such as pressure head measurements) while underestimates the efficiency of smaller numerical values (such as water content measurements). The efficiency is higher for the dataset composed of underestimated random values compared to the dataset composed of overestimated random values (see Random_5 and Random_6)

- The Coefficient of Variation (COV) and the normalized Coefficient of Variation (nCOV) show consistent and logic behaviour for the water content and pressure head datasets, they increase with the standard deviation of the input dataset. CV shows an irregularity in the mixed dataset (water content+pressure head).

- The KGE index shows an inconsistency in the water content dataset. KGE of Random_2 has a better fit (closer to 1) than Random_1 that is not logical because random_1 has a lower dispersion of the data compared to Random_2. KGE has a wider range of (erroneous) variation for the pressure head dataset in the positive and negative range with respect to pressure head+water content dataset that has minor fluctuations.
  The Kling-Gupta efficiency is born as improvement of the widely used NSE because it is based on an improved combination of the diagnostically meaningful components of the mean square error (mean, variability and dynamics) that are at the base of the NSE index (Murphy, 1988, Gupta et al, 2009). For our case study, NSE and its variants (NSEj and NSErel) seem better statistical indices than the evolution KGE.

- The index of agreement IA and its derivatives (IA_{rel}, IA_j) show an anomalous behaviour for all the three typologies of dataset. Moreover for the pressure head dataset the behaviour is also reversed (it decreases with good simulated values), while for water content and water content+pressure head datasets the direction is correct but not proportional to the data dispersion from the observed values (Random_2 has a value closer to 1 than Random_1). The index of agreement IA has been proposed to overcome the insensitivity of NSE and R^2 to differences in the observed and simulated means and variances but it results oversensitive to extreme values due to the squaring of the differences in the numerator (Willmott, 1981). As positive factors, IA is a relative and bounded measure that is widely used in every field for cross-comparisons between models/simulations. The performance of two indices ($IA_{j=1}$ and IA') introduced to overcome the limits of IA, has been tested. The first index was introduced by Legates and McCabe (1999) as a modification of IA [equation 124]:
- $$IA_{j=1} = 1. - \frac{\Sigma|M_i - S_i|}{\Sigma(|S_i - \overline{M}| + |M_i - \overline{M}|)}$$ [equation 124]

The modified index $IA_{j=1}$ gives an appropriate weighting to errors and differences, with no inflation given by squaring the values, one of the major limits of the original version of the coefficient of agreement IA. The range of variation of the index is 0-1 and 1 represents the best fit. The second modified index IA' (see equation 125) removes the influence of the simulated values from the denominator of IA, substituting them with the observed values. In this way the new denominator becomes a model-independent standard of comparison for the sum of errors (numerator).

$$IA' = 1. - \frac{\sum|M_i - S_i|}{2 * \sum|M_i - \bar{M}|} \qquad \text{[equation 125]}$$

The major downside of the index IA' is that it is unbounded on the lower end, so the interpretation of the model performance becomes more difficult. The application to the random dataset of the two indices modified from the original IA metric, shows a consistent and logical trend for IA' in all the dataset typologies, while for $IA_{j=1}$ the general trend is reversed (closer to 1 for wider dispersion of the data). IA' will be preferred for future analysis over the IA, IAj with j≥1; IA_{rel} indices.

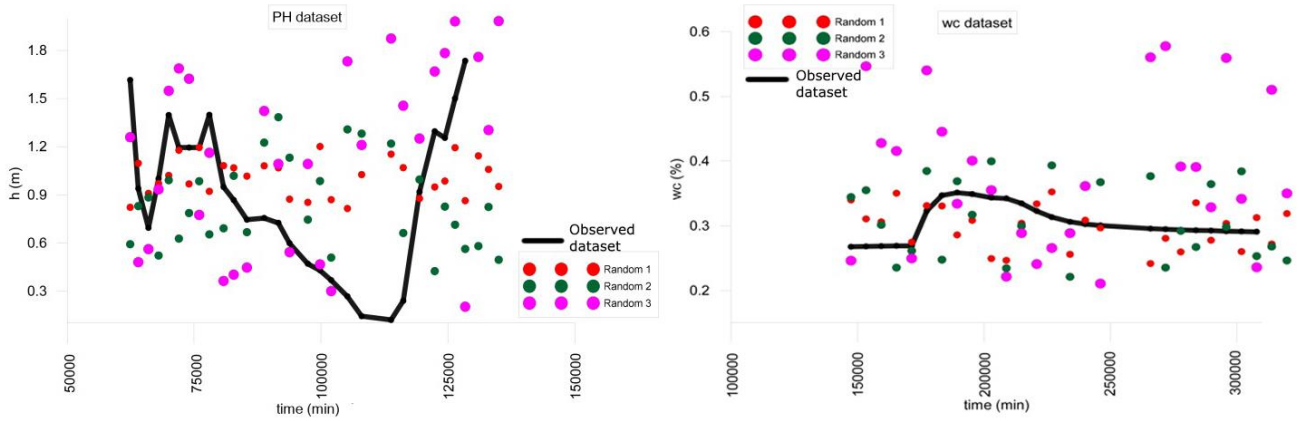


Figure 7 Graphs showing the dispersion of the random dataset of PH (left side) and water content (right side) around the observed dataset (black line).

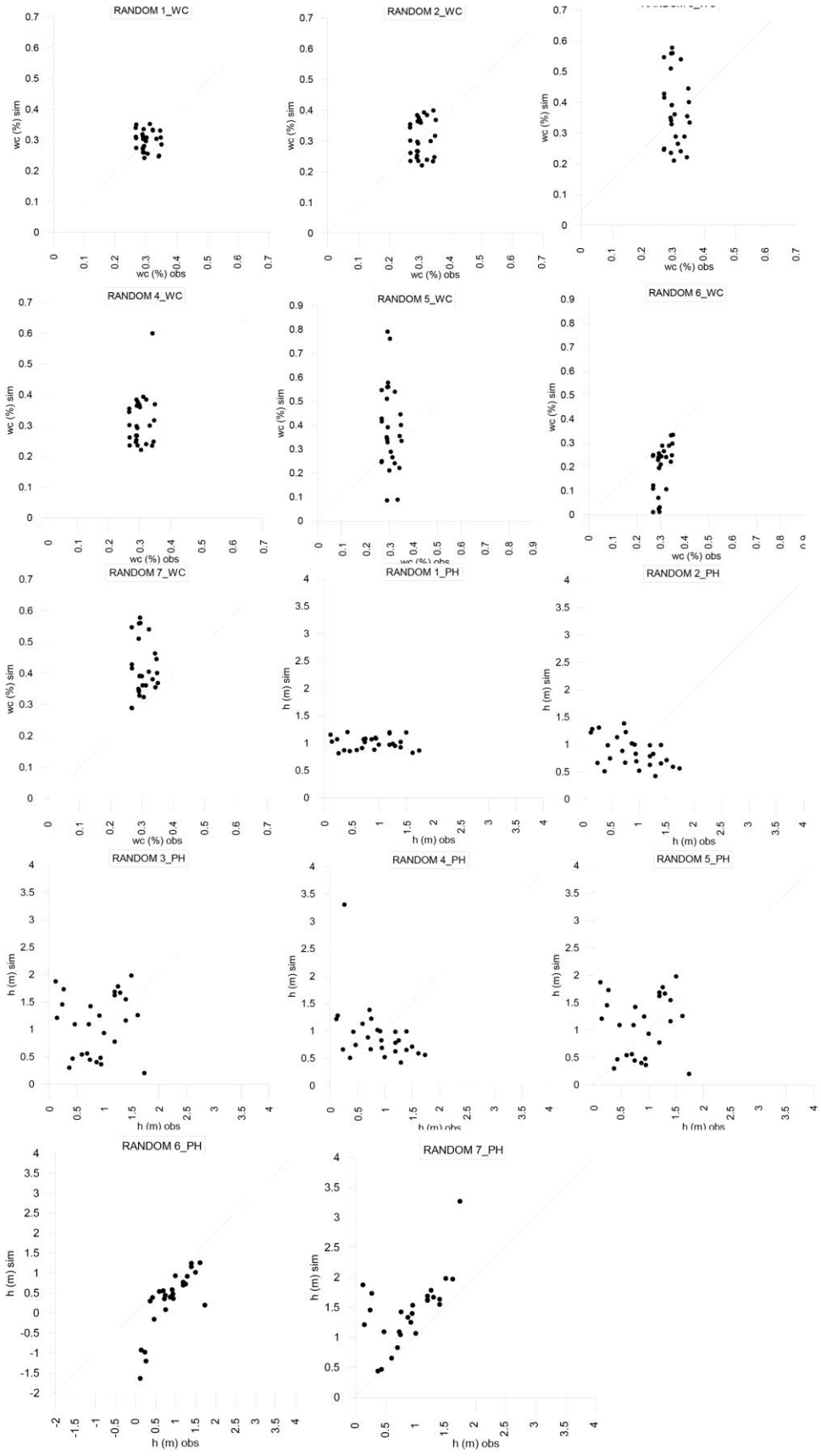


Figure 8 Graphs showing the dispersion of the simulated random datasets compared to the observed datasets. The more the points are allocated close to the 1:1 line, smaller is the dispersion.

Table 4: Metrics/indices values calculated for the different simulated random datasets-observed datasets. In red the metrics whose values are illogical with respect to the expected trend.

	RAND1_WC	RAND2_WC	RAND3_WC	RAND4_WC	RAND5_WC	RAND6_WC	RAND7_WC
MAE	0,04	0,05	0,11	0,06	0,15	0,11	0,11
MBE	0,01	0,00	-0,07	-0,01	-0,09	0,11	-0,11
MSE	0,00	0,00	0,02	0,01	0,04	0,02	0,02
RMSE	0,04	0,06	0,14	0,08	0,20	0,14	0,14
nCOV	2,14	3,87	7,48	5,20	11,35	6,37	5,53
COV	10,89	19,74	38,12	26,52	57,88	32,50	28,18
NSE	-1,88	-4,55	-27,08	-7,91	-56,25	-26,47	-27,08
IA	0,32	0,40	0,15	0,38	0,09	0,27	0,20
NSEj	-0,65	-1,47	-3,94	-1,81	-5,89	-3,90	-3,94
IAJ	0,27	0,28	0,15	0,25	0,12	0,20	0,18
NSE rel	-1,84	-4,50	-29,47	-7,13	-59,32	-28,68	-29,47
IA rel	0,33	0,40	0,08	0,43	0,04	0,21	0,13
CRM	0,02	-0,01	-0,19	-0,04	-0,22	0,55	-0,26
R^2	0,02	0,01	0,02	0,05	0,03	0,27	0,00
NSE log	-1,98	-4,59	-20,75	-6,64	-41,89	-47,29	-20,15
KGE	0,09	-0,53	-2,44	-1,16	-4,64	-1,75	-1,44
R	0,12	0,11	0,15	0,22	0,17	0,52	0,01
$IA_{j=1}$	0,27	0,28	0,15	0,25	0,12	0,20	0,18
IA'	0,18	-0,24	-1,47	-0,41	-2,44	-1,45	-1,47
	RAND1_PH	RAND2_PH	RAND3_PH	RAND4_PH	RAND5_PH	RAND6_PH	RAND7_PH
MAE	0,37	0,50	0,57	0,57	0,70	0,57	0,57
MBE	-0,12	0,05	-0,26	-0,02	-0,39	0,57	-0,57
MSE	0,22	0,36	0,53	0,64	1,02	0,53	0,53
RMSE	0,47	0,60	0,73	0,80	1,01	0,73	0,73
nCOV	2,48	5,65	11,87	10,91	17,73	15,02	11,38
COV	13,37	30,45	63,93	58,77	95,50	80,91	61,26
NSE	-0,17	-0,97	-1,86	-2,44	-4,50	-1,86	-1,86
IA	0,29	0,17	0,46	0,11	0,31	0,67	0,61
NSEj	-0,06	-0,43	-0,63	-0,62	-1,01	-0,63	-0,63
IAJ	0,24	0,15	0,36	0,13	0,32	0,43	0,43
NSE rel	-18,36	-23,56	-46,61	-39,92	-49,87	-46,61	-46,61
IA rel	-10,68	-9,33	-8,01	-9,64	-5,42	-4,51	-5,48
CRM	-0,12	0,06	-0,23	-0,02	-0,31	1,81	-0,39
R^2	0,00	0,22	0,00	0,19	0,00	0,62	0,33
NSE log	-0,23	-0,79	-1,35	-1,22	-1,80	-0,83	-0,91
KGE	-0,21	0,34	-0,02	0,40	-0,43	0,07	0,19
R	0,04	0,47	0,07	0,43	0,02	0,79	0,58
$IA_{j=1}$	0,24	0,15	0,36	0,13	0,32	0,43	0,43
IA'	0,47	0,29	0,19	0,19	0,00	0,19	0,19
	RAND1_PH+WC	RAND2_PH+WC	RAND3_PH+WC	RAND4_PH+WC	RAND5_PH+WC		
MAE	0,21	0,29	0,35	0,33	0,44		
MBE	-0,06	0,02	-0,17	-0,01	-0,25		
MSE	0,12	0,19	0,29	0,34	0,56		
RMSE	0,34	0,44	0,54	0,58	0,75		
nCOV	8,07	7,27	12,45	10,53	16,75		
COV	60,41	54,41	93,19	78,78	125,37		
NSE	0,37	-0,06	-0,58	-0,85	-2,04		
IA	0,80	0,59	0,69	0,45	0,56		
NSEj	0,42	0,22	0,05	0,11	-0,20		
IAJ	0,71	0,56	0,57	0,52	0,51		
NSE rel	-3,93	-5,27	-11,30	-9,44	-12,34		
IA rel	-0,53	-1,41	-1,41	-2,09	-0,94		
CRM	-0,09	0,04	-0,22	-0,02	-0,29		
R2	0,43	0,12	0,25	0,03	0,16		
NSE log	0,36	0,07	-0,30	-0,16	-0,64		
KGE	0,61	0,31	0,34	0,17	-0,06		
R	0,65	0,35	0,50	0,18	0,41		
$IA_{j=1}$	0,71	0,56	0,57	0,52	0,51		
IA'	0,71	0,61	0,52	0,56	0,40		

4_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The performance of different indices (accuracy metrics/efficiency criteria) has been tested prior to their application to the observed-simulated data of the investigated embankment section in order to individuate limits and potentialities of these indices when applied to different datasets. Random distributions of a simulated PH dataset, of a wc dataset and of a PH+wc dataset have been generated with an increasing dispersion of the data, or with an increasing number of outliers or with a global overestimation and underestimation of the simulated data with respect to the observed data. It has been observed that MAE, which is the simplest measure of forecast, shows a coherent trend in all the considered simulated datasets. It has to be remembered that this index is unbounded (from 0 to infinite) and dependent on the unit of measurement and for this reason a normalization is needed with mixed datasets. The same conclusions are found for MSE and its square root RMSE, logical trends are observed in all the investigated cases and both could be considered excellent metrics for numerical predictions. More difficult is the interpretation of MBE that could not be used to evaluate the goodness of a match between observed-simulated datasets but it gives only information on the general direction of the error. As observed, MBE fails to capture the increase in the simulated dataset dispersion for all the three types of considered data (PH, wc, PH+wc). The Coefficient of Residual Mass (CRM) gives a measure of the tendency of the model to underestimate or overestimate the observed data. It is an unbounded (-infinite to infinite) index and independent from the unit of measurement. It shows for the datasets wc and PH+wc illogical oscillations passing from the random dispersion 2 to random dispersion 3, while it shows a logical behaviour with an increase in the number of outliers.

Differently from MAE, MSE, RMSE, MBE and CRM, the coefficient of determination R^2 is bounded (0-1) and independent from the unit of measurement. It shows illogical oscillations with increasing dispersion of the data (see RAND1, RAND2, RAND3 in Table 4) which is in line with the limits of this index which could be greatly affected by outliers, by the number of data in the dataset and it could give high values (close to 1) even if the goodness of the model prediction is not sufficient due to, for example, problems of overfitting data (a model with too many variables compared to the number of observation points). It has been observed also the contrary, low values of R^2 even if model predictions are sufficiently good. This condition could happen if the residuals are widely dispersed and variance (σ^2) is high, R-Squared will inevitably be smaller but the regression line may still be the best way to describe the relationship between variables. For this reasons, it is always suggested to use this index together with others.

nCOV and COV (normalized coefficient of variation and coefficient of variation, respectively) show logical and consistent trend in all the considered datasets a part from the mixed dataset (wc+PH). It is a good indicator, not dependent on the unit of measurement but unbounded (0-infinite); it could give misleading results in case the observed-simulated datasets contain both positive and negative values and the mean of the data is close to zero. The index of agreement IA and its derivatives IA', IAj, IArel are unbounded indices, independent of the unit of measurement but they show an unsatisfactory behaviour for all the three considered datasets (RAND1, RAND2, RAND3). The reason could stay in the oversensitivity of these indices to extreme values. The derivate indices IA', $IA_{j=1}$ have been proposed to overcome this limit and IA' has shown a good behaviour in all the considered casuistries. IA' index will be preferred over the others in the followings. The Nash-Sutcliffe model efficiency coefficient (NSE) and its derivatives are good indicators, unbounded (1 to –infinite) and independent from the unit of measurement. They show, in all the considered cases, the logical trend we expect. NSE and its derivatives have oversensitivity to high values, moreover large values of NSE could be obtained with a poor model if data have high variability, while on

the contrary, if observations exhibit less variability and the values are close to the mean, NSE could approach negative infinity even if the model could predict well the observations. In order to address the shortcomings of NSE, the Kling-Gupta efficiency index (KGE) has been introduced. Despite this, in the performed tests, it has been observed an illogical behaviour of this index for the datasets of typology wc and PH+wc. For RAND2 and RAND3 in Table 4, KGE increases as the dispersion of the observed dataset increases (best fit 1) which is conceptually wrong. For this reason NSE and its derivatives will be preferred in the following analysis over KGE.

Due to the fact that each metric/index has its pros and cons and it is suitable to emphasize particular aspects of a simulated behaviour, it is always suggested to use a pool of metrics / indices on which the modeller is confident in order to have a complete overview on the simulation outcomes, overcoming possible misinterpretations due to specific limits of the single indices. Moreover, it is always suggested to use a limited number of indices/metrics in order to avoid to be lost in a large amount of data unable to draw any conclusions on the goodness of a model prediction. The metrics/indices that have shown in this chapter the expected behaviour for all the different typologies of random datasets will be preferred over the others in the following analysis.

Chapter §5 focuses on the elaborated numerical model of the 2018 simulation period. Geometry, initial conditions, boundary conditions, boundary fluxes, model discretization, iteration criteria have been explicitly dealt. A sensitivity analysis of the hydraulic parameters of the adopted hydraulic models (hysteretic and non-hysteretic ones) has been performed. Moreover the sensitivity of the model equations and solver has been tested considering different values of the input parameters with different number of digits after the comma in order to understand to which precision the problem has to be dealt (precision of the set of optimized parameters obtained by inverse analysis).

5 THE NUMERICAL MODEL OF THE INVESTIGATED EMBANKMENT SECTION

The 2D geometry of the investigated embankment of Secchia River, obtained by topographical survey, has been imported in the code by means of 47 geometrical points. The topographical survey investigated 40 m of section, crossing 3 times the embankment. The external outline used for the 2D numerical model is calculated as the average profile of the three sections. In Figure 9 the graphic result of the topographic survey is presented.

The layers in which the domain has been divided using CPTU tests interpretation are the ones presented in chapter §3. The resulting domain is 90 m wide from the centre of the river bed to the agrarian field in the outer-slope of the embankment and 23,4 m high. The origin of the geometrical reference system is in the centre of the river bed (x-axis) and 10 m a.s.l (y-axis). The 21 observation points are located in the positions of the installed instruments that are projected to the same 2D plain. The applied boundary conditions are summarised in the following Figure 9.

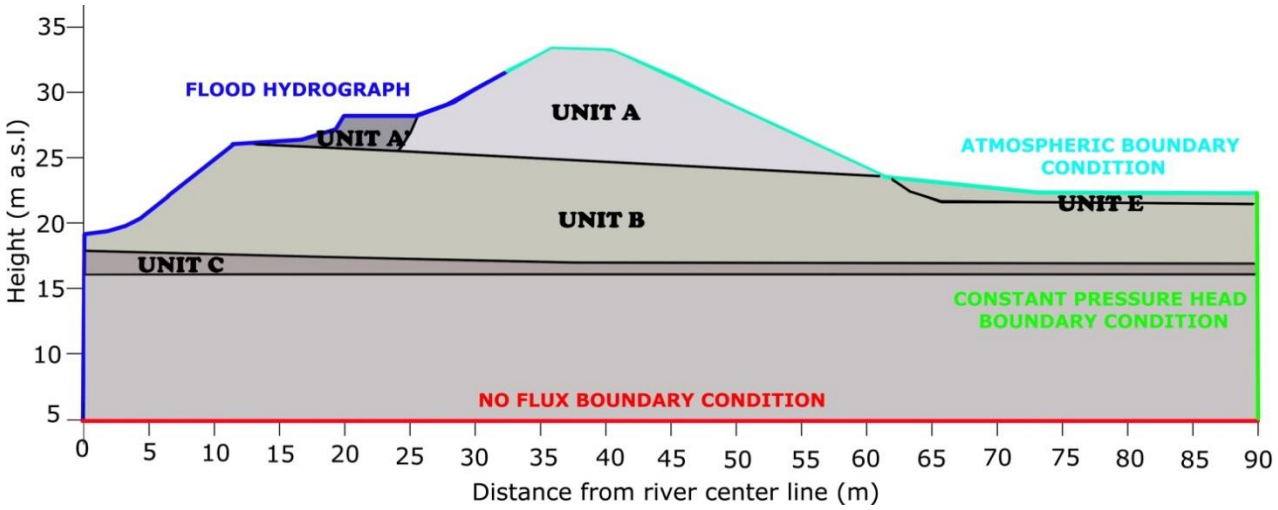


Figure 9: The geometry of the investigated section with indication of the soil layers and of the boundary conditions of the FE model.

The upper boundary is an atmospheric boundary condition, a ''system dependent boundary condition'' because the pressure head or the gradient along that boundary is not known a priori. It depends on the interaction between the soil and the environment for example the water flux across an atmospheric boundary conditions during a precipitation event depends on the precipitation rate and on the infiltration capacity of the superficial soil (Radcliffe and Simunek, 2010). To define the atmospheric boundary condition, data of precipitation, evaporation and transpiration every 30 min have been used. The investigated area lacks of a weather station. Data of humidity, temperature and precipitation are recorded by a weather station in Cortile da Carpi (Lat. 44.778387, Lon. 10.971285), 7 km away from the investigated section. These atmospheric information are available and freely downloadable from ARPAE (Regional Environmental Protection Agency) website, Dex3r web app (https://simc.arpae.it/dext3r/). To compute evaporation and transpiration contribution, CRITERIA 1D, an open-source software for soil water balance and crop modelling, developed by ARPA, has been used. Criteria uses a simplified model, assuming a multi-layered soil and computing explicitly values of water flows between layers, deep drainage, actual evapotranspiration flux, surface and subsurface runoff and capillary rise, following the approach of Driessen (1986) and Driessen and Konijn (1992). The model requires as input daily data of temperature (min-max), precipitation, the division in layers and for each layer the soil texture, bulk density, organic content, soil water retention curve and crop management information (sowing, tillage, fertilization, irrigation). If available, estimated potential evapo-transpiration could be given as input data. The choice to use a simplified and conceptual model as CRITERIA 1D is driven in our case by the lack of data to compute more precisely the actual evapo-transpiration contribution. In fact lysimeter data to measure the amount of actual evapo-transpiration in the investigated section are not available. Criteria 1D uses the Penman-Monteith (Penman,1949; Monteith, 1965) method, in the last revision by Allen et al (1994) to compute ETo, the Potential Daily Reference Evapo-transpiration, defined as the rate at which readily available water is vaporised from the reference vegetated surface. The reference surface is a uniform, dense grass of *festuca arundinacea* with a specific height and surface resistance in a field of at least 100 m of the same vegetation (Allen et al, 2005). Eto is a parameter independent of the type of crop that could be computed only from meteorological data. Penman-Monteith method is adopted and recommended by FAO and it is worldwide used as the reference. According to Penman-Monteith, ETo could be computed with the following expression [equation 126]:

$$ETO_{P\&M} = \frac{0{,}408\Delta(R_n - G) + \gamma \frac{900}{T_{average}+273} V_{wind}\,(e_s - e_a)}{\Delta + \gamma(1 + 0{,}34 * V_{wind})}$$ [equation 126]

where R_n is the net radiation ($MJm^{-2}d^{-1}$), G the net heat flux from the soil ($MJ\ m^{-2}\ d^{-1}$), Δ the slope of the saturated vapour function (kPa $°C^{-1}$), γ is the psychrometric constant (kPa $°C^{-1}$), V_{wind} is the average daily speed of the wind at a height of 2 meters (m/sec), e_s is the saturation vapour pressure (kPa), e_a is the actual vapour pressure (kPa). Average daily temperature, minimal and maximal daily relative humidity, solar radiation, wind speed are the input data for the Criteria software, all available in Dex3r webapp. Then to compute the maximum evaporation and the maximum transpiration the Kc, crop coefficient, is introduced. Kc is dependent on LAI, the Leaf Area Index. Criteria has provided a database of crop parameters for different type of culture; values available in literature. The description of crop parameters in the most precise and accurate way is surely recommended to increase the model accuracy e.g. computing LAI by means of a multi-spectral camera (Bittelli et al, 2010). Due to the lack of crop parameters determination for the investigated section, fallow is chosen from CRITERIA's database. The boundary condition between the embankment and the river (left side of the model) is a system-independent boundary condition, a Dirichlet boundary condition. The water level sequence through the investigated section is given in terms of pressure head (m) in time. The hydrometric data used in the numerical modelling are collected from the stream gauge at Ponte Bacchiello (Lat. 44.747546, Lon. 10.98734). A hydraulic 2D unsteady flow model using HEC-RAS (Hydrologyc Engineering Center's River Analysis System) is applied to correlate the water flow through Ponte Bacchiello (x-axis) with the water surface elevation (m) in Ponte Motta (y-axis). The conversion equation is the following [equation 127]:

$$y = -3{,}3678E\text{-}16x^6 + 9{,}6465E\text{-}13x^5 - 1{,}1202E\text{-}09x^4 + 6{,}8493E\text{-}07x^3 - 2{,}4196E\text{-}04x^2 + 5{,}6710E\text{-}02x + 2{,}2528E+01$$

[equation 127]

The lower edge of the model is assumed to be an impermeable layer, a system independent boundary condition in which the flux through is zero. To the right edge of the model a constant head of h=11 m is assigned to represent the hydrostatic conditions of the far-field water table. In order to create a robust seepage model, the various integration settings that could influence the results were tested prior to the modelling. In particular, to test the optimal mesh size, a line passing through the centre of the embankment is considered and the differences in the model output are investigated. The goal is to identify the minimum mesh size to obtain an accurate model without increase the computational burden. It is necessary to make a compromise because if for the embankment part subjected to higher hydraulic gradient a finer mesh is desirable, in the model extremities were the conditions are close to hydrostatic a finer mesh doesn't give any improvements to the model. In Figure 10 the pressure head profiles obtained for a section passing through the centre of the embankment in two different temporal instants (14610min (22/11/2017) and 43830 min (12/12/2017)) are compared using different mesh size (2m / 1,5m / 1m / 0,6m/ 0,3m) in order to individuate the best solution. For the mesh size 2,0m the line is extremely edgy because of few data points while for the mesh size below 1m the line is smooth and converge towards the ''more accurate'' solution. It was not possible to obtain convergence of the model for a mesh size of 0,3m and finer. Being the simulation to be performed on a medium-long period and being the number of inverse analysis to be simulated very high, a mesh size of 1 m has been considered for the ease of convergence and the reduced time of calculus. The mesh has been kept the same in all the performed simulations with the different hydraulic models (as will be described in the following). The mesh of the embankment core (unit A) and the fluvial deposit (unit A'), where the majority of the sensors are installed, have been refined to 0,3 m

together with the atmospheric boundary where evapo-transpiration phenomena and rainfall infiltration have to be computed. The chosen mesh is unstructured and it uses a network of triangular elements. The resulting mesh is composed of 4788 nodes, 685 1D elements due to the discretisation of boundary and internal curves and 9231 2D triangular elements. In Figure 11 the final mesh of the investigated section is shown.

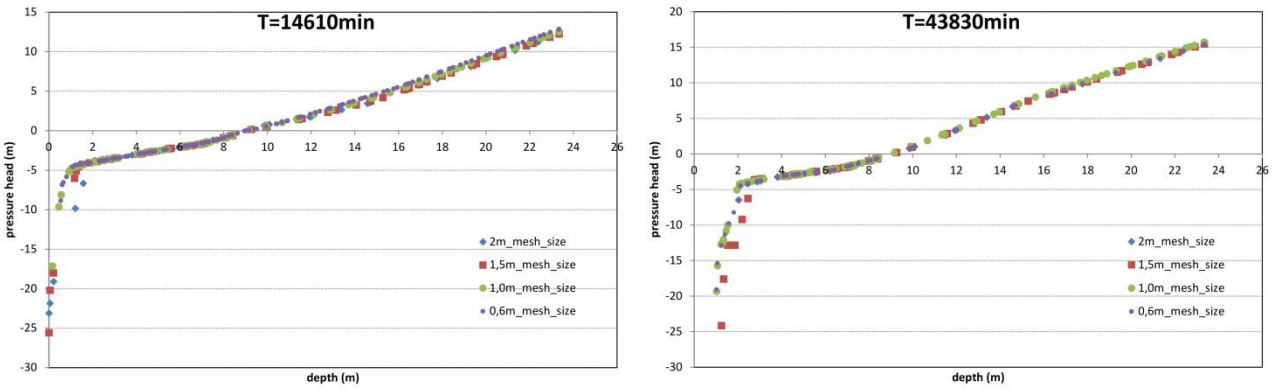


Figure 10: Pressure head profile of the section passing through the centre of the embankment (see Figure 11), in two temporal instants 14610 min (22/11/2017) and 43830 min (12/12/2017). Minutes are calculated from the initial time instant of the 2018 simulation (12th November 2017), see Figure 14.

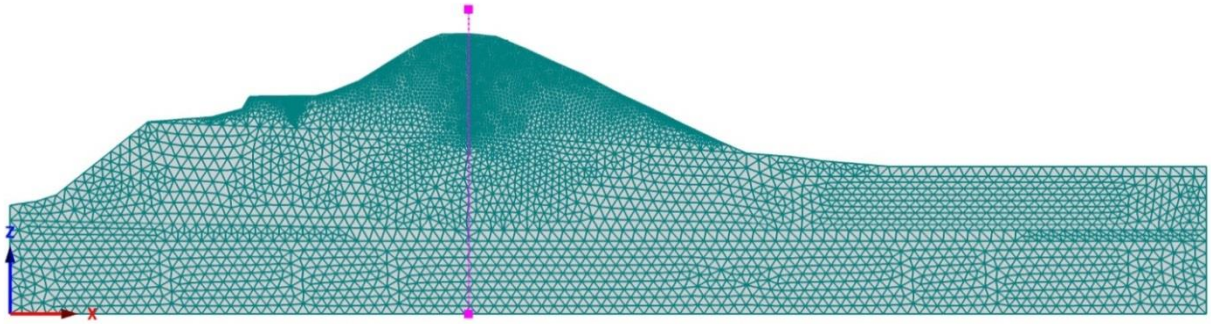


Figure 11: FE-Mesh used for the numerical modelling of the investigated section.

A similar ''model testing'' has been carried out to evaluate the best domain dimension. To do so, three different lengths of the right edge of the model are compared (x length =90, 100 and 110 m from the axis origin) through the solution in terms of pressure head calculated in three chosen sections (a,b,c) and in two temporal instants during high flood events: T=43830 min (12/12/2017) and T=180 000 min (16/03/2018), see Figure 14. In Figure 12 the location of the sections and in Figure 13 the compared pressure head profiles for the three configurations of the geometrical space are shown. As it is possible to observe, the configuration with 24,5 m of agricultural field on the right side of the embankment (x length from the axis origin equal to 100 m) is considered acceptable, being a very good compromise because, as could be observed in Figure 13, the pressure head profiles of geometries B and C are almost overlapped.

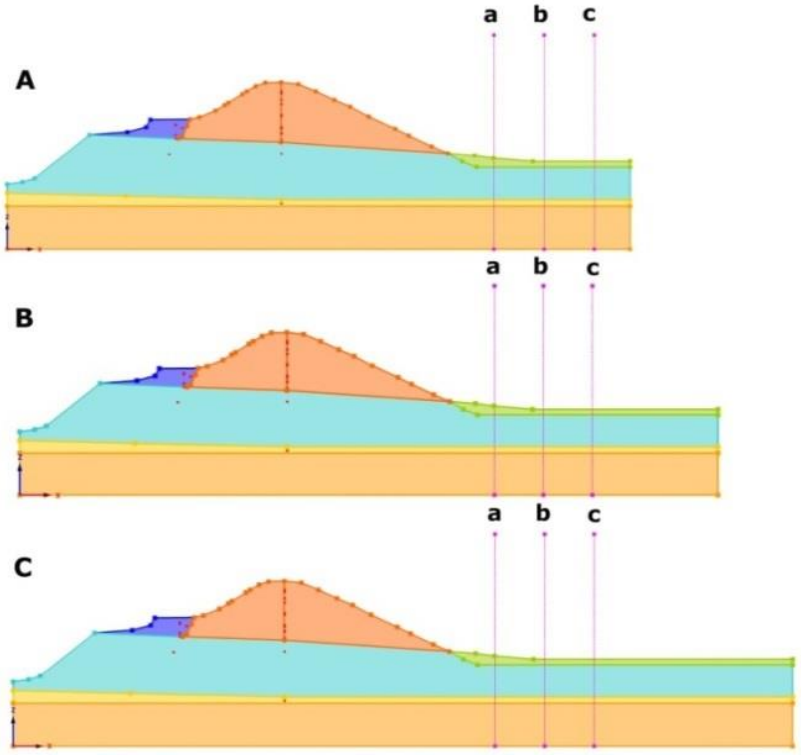


Figure 12: Position of the considered sections (a,b,c) in the different hypothesis of domain dimension (x max=90m (A),100m (B),110m (C)).

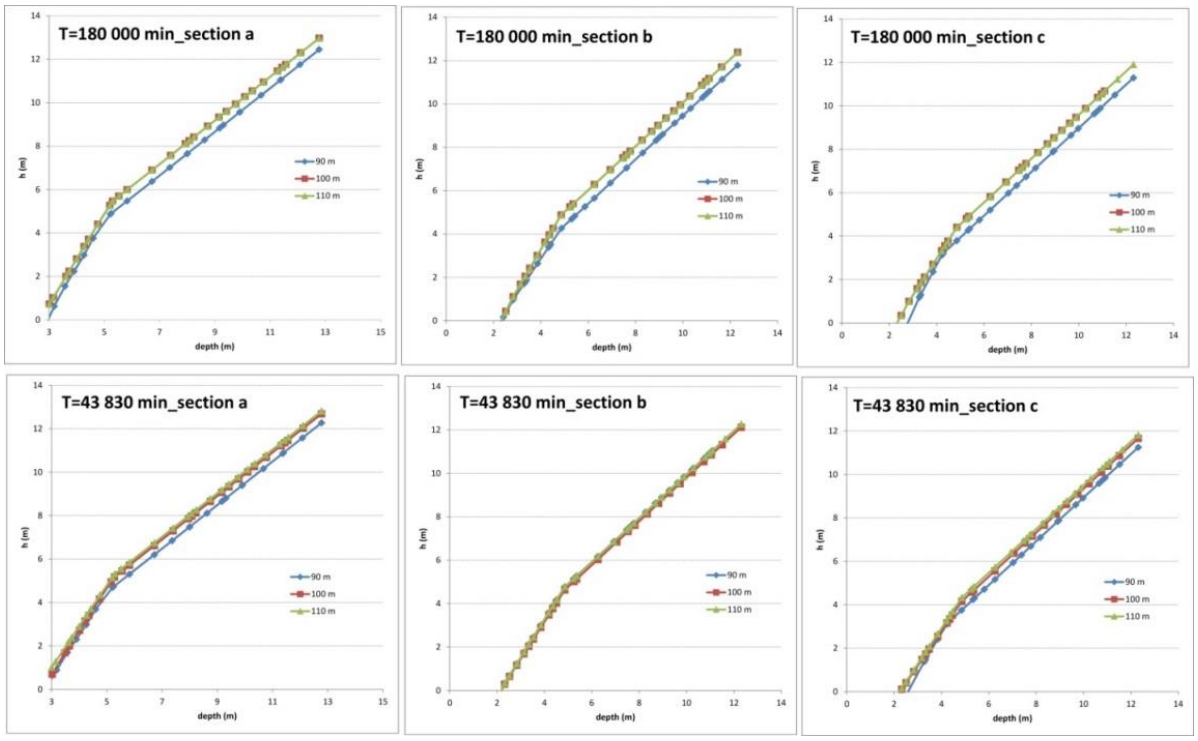


Figure 13: Pressure head profiles of sections a,b,c for three different geometrical dimensions of the domain (90,100,110 m from the axis origin) and for two temporal instants (180 000 min (16/03/2018) and 43830min (12/12/2017)). Minutes are calculated from the initial time instant of the 2018 simulation (12th November 2017), see Figure 14.

Regarding the analytical models for the retention curve parametrisation, four models have been chosen with which to run the direct and indirect simulations:

- ''VGM'': the van Genuchten model to describe the SWRC together with the statistical pore size distribution of Mualem (1976) to describe the hydraulic conductivity function K(h).
- ''HVGM'': the van Genuchten-Mualem model incorporating hysteresis in the SWCR and K(h) functions.
- ''VGM-2cm'': the van Genuchten model with a -2 cm air entry option.
- ''HVGM-2cm'': the van Genuchten-Mualem model incorporating hysteresis and the -2cm air entry option.

The adopted hydraulic models have been presented in great detail in chapter §2_3, §2_4 and §2_5 of the present thesis. Further in the thesis discussion, the van Genuchten model will be indicated with the acronym ''VGM'', the ''van Genuchten model considering -2cm air-entry value ''as ''VGM-2cm'', the ''van Genuchten model considering the hysteretic behaviour'' as ''HVGM'' and the ''van Genuchten model considering the -2cm air entry value and the hysteretic behaviour'' has ''HVGM-2cm''.

Regarding the integration time step, Hydrus-2D is always able to select the optimum time step starting from an initial time step, a minimum allowed time step and a maximum time step. While the maximum time step (maximum permitted value of time increment) has a very low influence on the modelling, the initial time step (initial time increment) and minimum time step (minimum permitted value of time increment) have a great influence, especially on convergence and calculation time. The initial time step has to be set to a large time step (5-15 min), in our case 5 minutes, as suggested by Simunek et al (2012) for long term process with variable boundary conditions. The minimum time step has to be set to a very low value, smaller than the initial time and smaller than the interval between time-variable conditions (in our case 30 min). For the investigated section this parameter turned out to be extremely relevant for the convergence of the model and its value was obtained with a long trial and error procedure. For the VGM, HVGM, VGM-2cm, the minimum time step has been set to 0,9 min, for the HVGM-2cm the minimum time step to obtain convergence has been set to 11 min.

The hydraulic parameters used for the simulation of period 2017 - 2018 are the average values of the parameters obtained from interpolation of laboratory data using the VG model. In chapter §3_2 the methodology with which laboratory data were obtained have been widely discussed and the hydraulic parameters obtained from each experimental SWRC presented in Table 3 together with a statistical analysis.

Initially, the modelled period has been the 2017, from the 30th January 2017 to the 11th November 2017. The 2017 period has been characterised by relatively low hydraulic levels of Secchia River (the only exception was the flood registered in the period 4th-7th February with a peak of 29,25 m a.s.l), and for this reason installed sensors didn't show a relevant answer to hydraulic stimuli. The initial conditions for the 2017 simulation have been identified using 5 information: two Casagrande piezometers in the foundation, T8-TB1 in the berm, the level of the groundwater below the agrarian field (1,5 m below the surface) identified by CPTU test and the hydrometric level of the river. Using these 5 information, the geometry of the groundwater table has been traced. Above the groundwater, a non-hydrostatic but linear negative pore pressure distribution has been adopted. For the central part of the embankment (Unit A), SPC1 (-7m from the crown, h value=-2,1 m) and MPC1 (-4,7 m from the crown, h value=-2,8m) have been used to trace the slope of the curve pressure head (x)-depth (y); while MPC1 (-3,05m, h value=-1,7 m) has been excluded from the interpolation because it is affected by the proximity to the atmospheric boundary. For the berm (Unit A'), MPB1 (0,9 m from the berm surface, h value= - 0,53 m) and MPB2 (2,7 m from the berm surface,

h value= -0,53 m) have been used to trace a linear interpolation of the pwp profile. Only a small part of the sensors was available or reliable (not yet in equilibrium with the surrounding medium) due to the fact that the installation campaign was not finished at the time and some of the sensors have been installed only recently.

The results in terms of water content and pressure head distribution of the last temporal instant of the 2017 period are used as imported initial conditions for the simulation of the 2018 period. The 2018 simulation covers the period from the 12th November 2017 to the 23rd June 2018 (see Figure 14).

For the 2017 period, a direct simulation has been performed using only the VGM while for the 2018 period, direct simulations have been performed using the four hydraulic models available in the Hydrus library (VGM, HVGM, VGM-2cm, HVGM-2cm). The calibration phase has been performed on the 2018 period which is more significant in terms of pore pressure/water content distribution changes due to a rapid succession of high water peaks which have followed each other throughout the whole period.

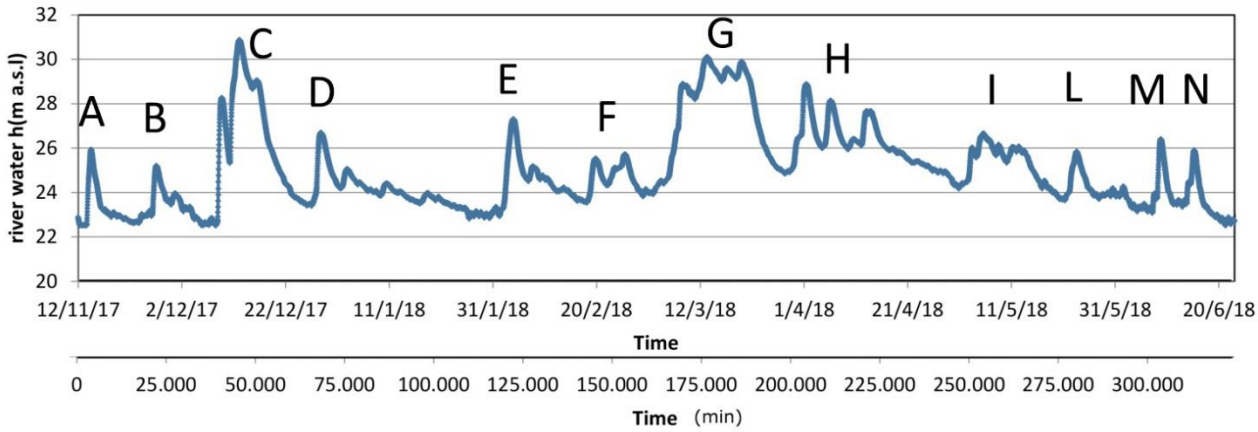


Figure 14: River water height of the Secchia River (hydrometer in Ponte Motta). On the x-axis the temporal scale in days and in minutes (used for the numerical modelling in Hydrus 2D). In the following the used time unit will be ''minutes'' from the initial instant of the simulation (12th November 2018) but, for sake of clarity, the day will be always reported.

5_1 SENSITIVITY ANALYSIS OF THE HYDRAULIC PARAMETERS

Sensitivity analysis assesses the effects on the output of a model induced by changes in the values of the input variables. Sensitivity analysis is used to define the most relevant parameters of the simulation, the parameters which have less influence and could be eventually eliminated from the model equations (reduction of the model), the parameters that make the output more susceptible to changes, high correlated parameters with the output (small changes in the parameters are able to produce big changes in the output). Moreover, an initial sensitivity analysis is essential to reduce the number of parameters in the optimization process by eliminating the insensible ones (Mashayekhi et al, 2016). It is relevant to underline that parameter uncertainty and sensibility are inversely proportional since highly-sensitive variables result in a small parameter uncertainties. To analyse sensitivity, the one factor a time (OFAT) technique could be used: in a system with k parameters, the value of one parameter is changed and the remaining k-1 parameters must be set to their base value. The set of base values takes the name of ''base case'' which corresponds to the best bet of the parameters. This technique is the most widely used and the simplest due to the reduced computational effort but it does not consider the simultaneous variation of the parameters and the influence on the output and it could not investigate the correlation between parameters. When

dealing with models with a great number of parameters, the sensitivity analysis is able to give indication on the sensible parameters on which the calibration should work, avoiding time consuming on low sensible parameters. To assess the sensitivity of the different hydraulic parameters, the sensitivity coefficient $S(t,b_j)$ calculated according to Abbasi (2015) and Simunek et al (1998) is applied (see equation 128). The coefficient is independent of the measure unit and it allows a direct comparison between parameters.

$$S(t, b_j) = \frac{|Y(b_j + \Delta b_j e_j) - Y(b_j)|}{Y(b_j)} \quad \text{[equation 128]}$$

Where $S(t,b_j)$ is the change in the variable Y corresponding to a 1% change in the parameter b_j; Δb=0,01 b, e_j is the j_{th} unit vector. The coefficient of sensitivity is able to investigate the behaviour of the objective function in the parameters space and, in particular, in vicinity of the base set of parameters. If a relevant sensitivity is detected, it means that the minimum of the objective function is well defined and it is possible to estimate the parameters with relatively high precision in the inverse problem. The coefficient is not able to give information about local minima and the behaviour of the objective function in other spots of the parameters space (Simunek et al, 2008) but it remains the easiest index to evaluate model sensitivity with a one-factor-at-a-time-method OFAT analysis.

For each layer, an observation point is considered, then starting from the base set of parameters one parameter at a time is varied of +/-1% of its value. For the embankment section the observation point MPS6-SPC1-7m is considered, for the fluvial layer GS3-MPB2-2,2m, for the foundation layer T8-TB1-4,9m and for the aquifer layer PZ1-17m. The sensitivity coefficients have been calculated in each considered observation point in terms of PH, in order to allow an easier comparison of the results between the different points. In Table 5 the base values of the parameters and the varied values of the parameters (+1% of the base value), used in the OFAT analysis, are presented.

In Figure 15, the sensitivity coefficient vs time for each layer and under the condition +1% of the initial value of the considered parameter is presented. It has been chosen to not present here graphs obtained under the condition -1% of the initial value of the considered parameter for reason of brevity.

Focusing the attention on the hydraulic parameters of the drying curve (θr, θs, n, Ks, α), the parameters with higher sensitivity for the embankment layer are ''n'' and ''θs''. Two peaks of sensitivity are well defined around 73470 min (2/01/2018) and 188200 min (22/03/2018), both are in correspondence of floods (flood D and G respectively in Figure 14). Flood G is particularly relevant (high water level and persistent in time). During the initial and final time period (before and after winter period), the parameters of the embankment layer show a very low sensitivity (close to 0). For what concerns the fluvial layer, peaks of sensitivity of n, θs and α are registered in correspondence of 50 000 min (16/12/2017) and 200 000 min (30/03/2018) (flood D and G/H in Figure 19), while in the remaining period the sensitivity is zero. The global sensitivity of the fluvial layer is lower than the one of the embankment layer. As we could expect, the sensitivity of the parameters of the aquifer layer is zero a part from the saturated permeability Ks, with an average sensitivity of 0,08. In fact the aquifer layer is saturated due to the presence of the water table, for this reason the only parameter that influences its behaviour is the saturated permeability. The foundation layer shows high sensitivity for the parameter ''n'' around 41800 min (11/12/2017, flood C) and 304200 min (11/06/2018, flood M in Figure 14), the same periods observed also for the embankment and fluvial layers, but with the difference that the sensitivity remains high during all the investigated period. The pore-connectivity parameter ''l'' has not been investigated due to previous studies stating the lowest sensitivity

of this parameter compared to the other hydraulic parameters (Abbasi et al, 2003). We can conclude that the calibration procedure to be carried out must focus on the optimization of parameters n, θ_s, Ks, α for the embankment layer, Ks for the aquifer layer, n, θ_s, θ_r, Ks, α for the fluvial layer and θ_r, n, θ_s, Ks for the foundation layer. It is important to notice that a variation of +1% of the initial value of the parameter causes relevant changes in the model output[1] and this aspect is extremely interesting for the investigated case study because a number of performed indirect simulations (as we will see in the following) have the optimized parameters that result in a narrow range close to the base-values, some of them with just changes in the fourth or fifth digit after the comma .The optimized parameters obtained from inverse simulation have a precision of 5 digits after the comma and the performed sensitivity analysis gives indication if it is relevant or not to maintain such precision. The parameters that we have underlined for their incidence on the model output have to be treated with the highest possible precision. The initial parameters dataset (base dataset) has a precision of 3 digits after the comma because a higher precision seems not reasonable considering that their values are the average of the hydraulic parameters obtained by laboratory SWRC.

Focusing the attention on the hydraulic parameters of the wetting curve (θ_m; θ_{sw};α_w; K_{sw}) in Figure 15, it is possible to observe a lower sensitivity compared to the hydraulic parameters of the drying curve. The average value of the sensitivity coefficient for the three layers (embankment, fluvial, foundation) is +/-0,02 while for the hydraulic parameters of the drying curve is around +/- 0,06. The parameters θ_m, θ_{sw} , α_w have the same trend in time for the three layers while K_w has a different trend. The embankment layer shows a peak of sensitivity around 180 000-200 000 min (flood G in Figure 14), while the fluvial layer around 60 000 min (flood C). Floods G and C are the major flooding events of the 2018 year. For what concerns the foundation layer peaks of the sensitivity coefficients are observable during the whole 2018 period, as observed for the parameters of the drying curve. The sensitivity analysis of the wetting parameters shows, as could be expected, the greatest influence on the model output during the most relevant flooding events (for the water level height or duration in time).

	EMBANKMENT		FLUVIAL		FOUNDATION	
VGM	**BASE VALUES**	**+ 1%**	**BASE VALUES**	**+ 1%**	**BASE VALUES**	**+ 1%**
α (1/m)	0,850	0,8585	0,250	0,2525	0,117	0,1181
θ_s	0,397	0,4009	0,319	0,3221	0,424	0,4282
K_d (m/min)	0,000090	0,0000909	0,0000276	0,0000304	0,000000198	0,000000218
n	1,304	1,3170	1,284	1,2968	1,14200	1,1534
θ_r	0,004	0,00404	0,001	0,00101	0,001	0,00101
HVGM						
θ_m	0,3975	0,4372	0,3191	0,3510	0,4241	0,4665
θ_{sw}	0,3573	0,3930	0,287	0,3157	0,3816	0,4197
α_w	1,70	1,87	0,500	0,550	0,240	0,264
K_{sw} (m/min)	0,000080	0,000088	0,000010	0,000011	0,00000010	0,00000011

Table 5:base values of the parameters for the embankment, fluvial and foundation layers and the varied values of the parameters (+1%of the initial value) used in the OFAT analysis.

[1] If we consider for example a sensitivity coefficient of 0,1 and a value Y(bj)=1 m, the difference between the output of the model with the varied parameter and the output of the model with the base parameter ($Y(b_j + \Delta b_j e_j) - Y(b_j)$) is equal to 0,1 m which is a significative number in a pwp distribution analysis.

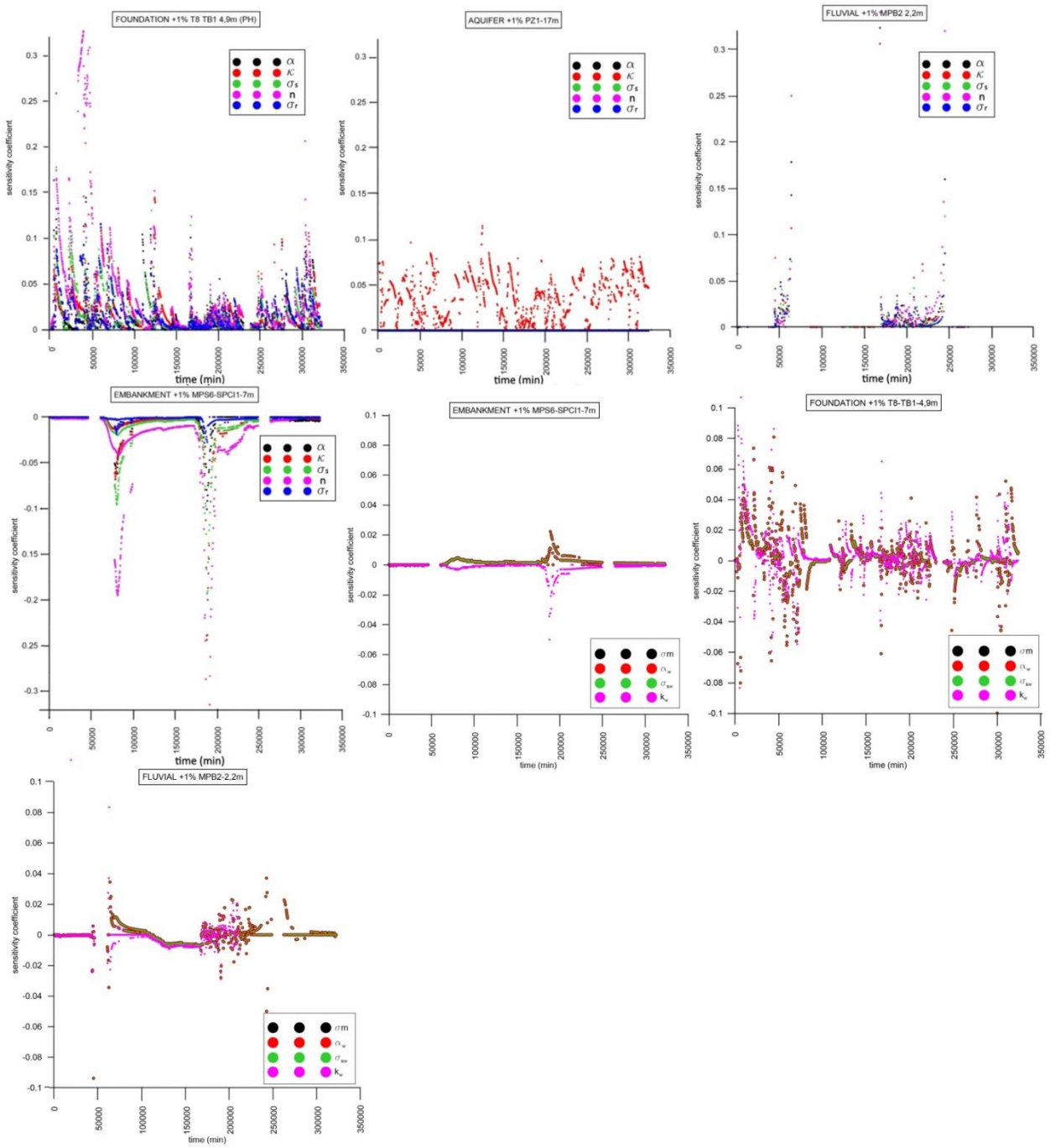


Figure 15: Sensitivity coefficient in time calculated in the chosen observation points (SPC1-7m for the embankment layer; PZ1-17m for the aquifer layer, MPB2-2,2m for the fluvial layer and TB1-4,9m for the foundation layer) changing a parameter at a time of +1% of its value (OFAT analysis). Minutes on the x-axis are calculated from the initial time instant of the 2018 simulation (12th November 2017), see Figure 14.

5_2 INVESTIGATION ON THE SIGNIFICANT DIGITS OF THE MODEL PARAMETERS

In Hydrus 2D the hydraulic input parameters could be set with a different number of digits after the decimal point: ''θs'' fifteen digits, ''θr'' eighteen digits, ''α'' fifteen digits, ''n'' fourteen digits and ''K'' thirteen digits. The output values in terms of observed-simulated values (PH or wc) have 3 digits after the comma and the optimized parameters in output (in the indirect problem) could have up to five digits after the comma. The set of base values (initial set of hydraulic parameters) used in the present case study has 3

digits after the comma and, as said before, it corresponds to the average values of the parameters as obtained by laboratory tests (evaporation tests).

The sensitivity of the model equations and solver has been tested considering different values of the input parameters with different number of digit after the comma (from three to five for θs, θr, α, n and from 5 to 11 for K) for the embankment, fluvial and foundation layers. One parameter at a time is changed and the differences between the output and the base solution in the observation points of the same layer are computed. Due to the fact that the optimized parameters obtained in the calibration phase (from the inverse analysis) have 5 digits after the point, it is important to know the degree of sensitivity of the model to changes of the parameter value after the third digit. We are interested to know if a direct simulation that shows a change in the fourth or fifth digit with respect to the base value is significant. This aspect is extremely interesting for the investigated case study because a number of performed indirect simulations give in output values of the optimized parameters that result in a narrow range close to the base-values, some of them with just changes in the fourth or fifth digit.

In Table 6 the values with which the simulations have been performed are reported. As it is possible to see, the tested values are very close to the base values but with an increasing number of digits after the comma. The observation points considered to compute the difference between output with base parameter and output with the changed parameter are:

-MPS6-MPC1-4,6m; MPS6-SPC1-7m; T8-TC2-8m for the embankment layer (pressure head measurements)

-GS3-MPC1-2,4m; GS3-MPC1-4,5m; GS3-SPC2-7,1m for the embankment layer (water content measurements)

-GS3-MPB2-2,2m for the fluvial layer (water content measurements)

-MPS6-SPB1-1,2m for the fluvial layer (pressure head measurements)

-T8-TB1, PZ2-10m for the foundation layer (pressure head measurements)

Table 6: The table reports the base values of each hydraulic parameter and the varied parameter with an increasing number of digits after the comma. These parameters are tested in order to investigate the significant number of digit to consider for each parameter.

EMBANKMENT									
	BASE VALUES	**TESTED VALUES**							
α (1/m)	0,850	α_1	0,8504	α_2					
θs (-)	0,397	θs_1		θs_2					
K (m/min)	0,00009	K_1	0,000094	K_2	0,0000944	K_3	0,00009444	K_4	0,00009444
n (-)	1,304	n_1	1,3044	n_2	1,30444				
θr (-)	0,004	θr_1	0,0044	θr_2	0,00444				
FLUVIAL									
α (1/m)	0,250	α_1		α_2					
θs (-)	0,319	θs_1		θs_2					
K (m/min)	0,00002	K_1	0,000024	K_2	0,0000244	K_3	0,00002444	K_4	0,000024444
n (-)	1,284	n_1	1,2845	n_2	1,28455				
θr (-)	0,001	θr_1	0,0015	θr_2	0,00155				
FOUNDATION									

α (1/m)	0,117	**α_1**	0,1174	**α_2**	0,11744				
θs (-)	0,424	**θs_1**	0,4244	**θs_2**	0,42444				
K (m/min)	0,0000001	**K_1**	1,4E-07	**K_2**	1,44E-07	**K_3**	1,444E-07	**K_4**	1,4444E-07
n (-)	1,142	**n_1**	1,1424	**n_2**	1,14244				
θr (-)	0,001	**θr_1**	0,0014	**θr_2**	0,00144				

For each simulation, the difference between the output of the model with the changed parameter and the output with the base value is computed. In the table below the maximum and minimum differences (in terms of PH or wc) for each performed simulation are presented.

Table 7: The table reports the differences in terms of PH (m), wc (-) between the simulation with the base parameter and the simulation with the varied parameter.

		α_1	**α_2**	**K_1**	**K_2**	**K_3**	**K_4**	**n_1**	**n_2**	**θr_1**	**θr_2**	**θs_1**	**θs_2**
PH EMB	**Max (m)**	0,009	0,004	0,153	0,174	0,016	0,016	0,010	0,009	0,006	0,015	0,016	0,016
	Min (m)	-0,005	-0,010	-0,112	-0,128	-0,023	-0,023	-0,018	-0,008	-0,007	-0,009	-0,023	-0,023
wc EMB	**Max (-)**	0,001	0,001	0,002	0,002	0,002	0,087	0,001	0,001	0,000	0,001	0,00	0,001
	Min (-)	0,000	0,000	-0,006	-0,008	-0,007	0,013	0,000	0,000	-0,001	-0,001	-0,001	-0,001
wc FLUV	**Max (-)**	0,001	0,001	0,002	0,002	0,002	0,002	0,001	0,001	0,001	0,000	0,000	0,000
	Min (-)	-0,001	-0,001	-0,024	-0,024	-0,024	-0,024	0,000	0,000	-0,001	-0,001	-0,001	-0,001
PH FLUV	**Max (m)**	0,118	0,120	0,148	0,166	0,167	0,168	0,171	0,171	0,170	0,171	0,037	0,037
	Min (m)	-0,074	-0,054	-0,316	-0,379	-0,354	-0,380	-0,380	-0,380	-0,381	-0,347	-0,039	-0,039
PH FOUND	**Max (m)**	0,331	0,282	0,318	0,384	0,379	0,417	0,417	0,224	0,581	0,227	0,213	0,164
	Min (m)	-0,207	-0,143	-1,758	-1,762	-1,741	-1,779	-1,778	-0,200	-0,235	-0,355	-0,198	-0,169

The bar diagram (tornado plot) is a powerful tool in deterministic sensitivity analysis to compare the relative importance of variables. It shows for each simulation the lower and higher value of the difference between base line output (output with base parameter) and the output with the changed parameter (α_1,α_2,K_1,K_2,K_3,K_4,n_1,n_2, θr_1,θr_2,θs_1,θs_2). It shows in an effective way the parameters that contribute the most to the variability of the outcome, reporting graphically the information provided in Table 7.

As it is possible to notice in Figure 16, the parameter K is the one that drives the largest impact on the output, changes of the decimeter order are reported for the pressure head dataset of the embankment (~0,2m) and fluvial layers (~0,4m) while changes of the meter order (~ 1,7m) for the foundation layer. For the water content dataset, variation of the K parameter produces changes of the order 0,002-0,087 (-) for the fluvial layer and 0,002-0,024 (-) for the foundation layer. Relevant changes are observable also for θs in the embankment layer (PH dataset) with maximum difference ~ 0,02m; for α in the fluvial layer (PH dataset) with maximum difference ~ 0,1 m and in the foundation layer (PH dataset) with maximum differences of 0,30 m and 0,60 m for the α and θr parameter, respectively.

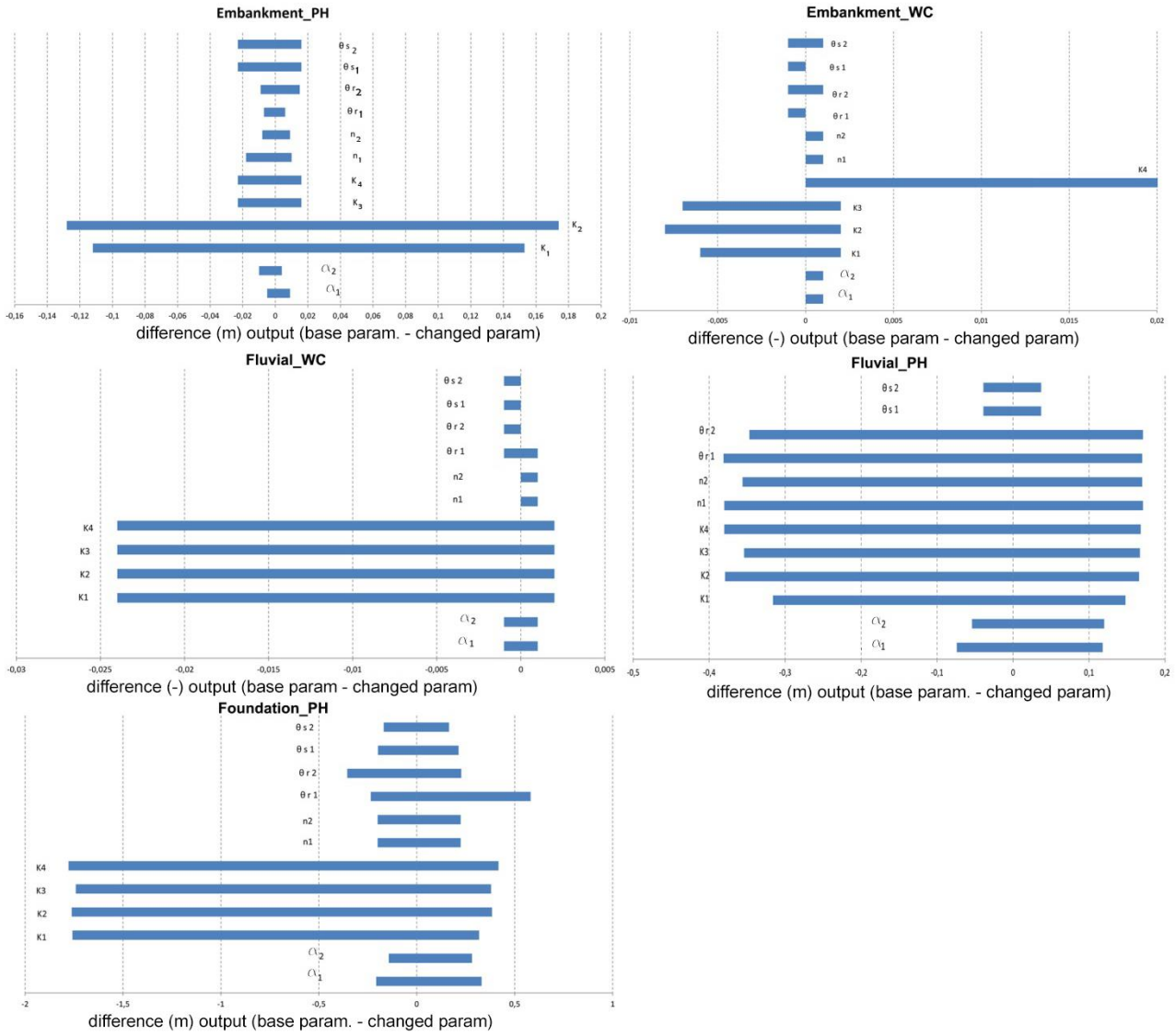


Figure 16: Tornado plot representing the max and min differences between the output of the simulation using the base parameter and the varied parameter (α_1,α_2,K_1,K_2,K_3,K_4,n_1,n_2, θr_1,θr_2,θs_1,θs_2).

An analytical method to assess the sensitivity of a system of parameters on the model output uses sensitivity indices. For the investigated case study the sensitivity index of Gomes et al (2017), developed from the work of Saraiva et al (2017), is proposed. The method is applied to system output values obtained by one-at-a-time-measures. The differences between output value and the base solution are calculated for each parameter of the system then a normalization is applied. The proposed method is a generalization of the method of sum of differences. The used sensitivity index has the following formulation:

$$S(x_i) = \frac{\frac{1}{n}\sum_{j=1}^{n}|y_{ij}-b|}{\sum_{i=1}^{k}(\frac{1}{n}\sum_{j=1}^{n}|y_{ij}-b|)} \qquad \text{[equation 129]}$$

Where "n" is the number of one-at-a-time measures of the parameter x_i, " i" is the index of the parameter, " y_{ij}" is the output of the system for the j-th measure of x_i, " k" is the total number of parameters of the model and " b" is the base solution. In Table 8 the sensitivity indices $S(x_i)$ are summarized for each family

of simulations (PH_embankment, wc_embankment, wc_fluvial, PH_fluvial, PH_foundation). The index is able to show numerically the impact on the model output of an increasing number of digit after the comma in the values of the adopted parameters. To help visualize this concept, a chromatic scale from yellow (lowest sensitivity) to orange (highest sensitivity) is applied in Table 8. As it is possible to notice the α parameter is the one that shows the lowest impact in each layer together with the θr parameter. On the contrary the saturated permeability K shows a relevant impact on the model output in fact changes even in the ninth-eleventh digit after the comma could produce non-negligible changes on the simulation results.

Table 8: The table reports the max value of the sensitivity coefficient found for each layer (embankment, fluvial, foundation) using a PH dataset or a wc dataset.

	SENSITIVITY INDEX											
	α_1	α_2	K_1	K_2	K_3	K_4	n_1	n_2	θr_1	θr_2	θs_1	θs_2
PH_emb (m)	0,007	0,009	0,334	0,382	0,051	0,051	0,025	0,011	0,009	0,018	0,051	0,051
wc_emb (-)	0,001	0,002	0,018	0,022	0,021	0,891	0,005	0,005	0,004	0,005	0,013	0,014
wc_fluv (-)	0,002	0,004	0,166	0,178	0,178	0,177	0,004	0,002	0,001	0,003	0,098	0,187
PH_fluv (m)	0,008	0,007	0,112	0,121	0,122	0,123	0,123	0,125	0,123	0,124	0,006	0,006
PH_found (m)	0,015	0,067	0,055	0,133	0,149	0,132	0,129	0,057	0,057	0,073	0,065	0,067

5_3 PRELIMINARY CONLUSIONS AND FURTHER STUDIES

The geometry of the embankment section along Secchia River has been investigated by means of a topographic survey, then the spatial points have been imported in the Hydrus2D program as coordinates x,y. The resulting domain is 90m wide and 23,4m high (see Figure 9). N°21 observation points have been positioned in the model section where sensors have been installed in the bank body. The upper boundary of the model section is an atmospheric boundary defined by data in time (every 30 min) of precipitation, evaporation and transpiration. The evapo-transpiration contribution has been computed using the free-source program CRITERIA 1D developed by ARPA, an algorithm which approximates physical processes through simplified schemes. This simple modelling approach is mandatory in the investigated case study due to the lack of the necessary parameters (to compute the actual evapo-transpiration contribution from the potential contribution elaborated using the Penman-Monteith method (Penman,1949; Monteith, 1965)) for a more accurate representation of the phenomenon. The lower boundary of the model is a no flux boundary, while on the left side the hydrographic level monitored by a stream gauge at Ponte Bacchiello is imposed by means of pressure head (m) data in time. To the right edge of the model, a constant head boundary condition is applied to represent the far-field water table. In order to individuate the best mesh size to discretize the model, the results in terms of pressure head in the section passing through the embankment crest have been compared using a mesh dimension from 2m to 0,3m. A dimension of 1m guarantees the almost perfect coincidence with the ‘‘more accurate’’ solution obtained with a mesh size of 0,3m, and, in close proximity to the observation points and close to the atmospheric boundary condition, the mesh has been refined up to 0,3m in order to ease the computation where the majority of the numerical problem complexities are expected (see Figure 11). The optimal horizontal dimension of the model has been tested in order to individuate the minimal extension of the right edge which guarantees the ‘‘more accurate solution” (see Figure 12; Figure 13). A dimension of 100 m from the axis origin (center of the river bed) has been individuated as the optimal one. Four different hydraulic models have been used to simulate the 2017-2019 hydrometric river water heights succession in the

investigated section: the Van Genutchen model (VGM), the Van Genutchen model considering an air entry value of -2cm (VGM-2cm), the Van Genutchen model considering the hysteretic behaviour (HVGM) and the van Genuchten-Mualem model incorporating hysteresis and the -2cm air entry value (HVGM-2cm). The adopted hydraulic parameters of the SWRC models are the average values of the parameters obtained by laboratory tests. The period from 30th January 2017 to 11th November 2017 has been simulated and calibrated in order to use the pwp distribution of the last instant of the simulation as initial conditions for the simulation of the 2018 period (12th November 2017 to 23rd June 2018). 2018 period has a greater relevance for the addressed geotechnical problem compared to the 2017 period due to the greater number of flood peaks registered and their persistence in time (see Figure 14). The initial conditions for the simulation of the 2017 period have been obtained using the monitoring data of the first day of simulation in the observation points and considering a linear interpolation of the pressure head values in the remaining points. A sensitivity analysis of the hydraulic parameters of the wetting and drying curves has been performed in order to individuate the most relevant ones on whom the calibration has to focus. The OFAT methodology has been applied (Simunek et al (1998), Abbasi (2015)): one hydraulic parameter at a time is varied of +/-1% with respect to the initial value and the sensitivity coefficient is measured in each layer of the model in order to investigate the local response to parameter change. Looking closely to the results increasing the value of the hydraulic parameters of +1% (with respect to the base value), it has been observed that n, θs of the embankment layer show the highest sensitivity with peaks in correspondence of floods D and G (see Figure 14) while in the remaining simulation period the sensitivity is low or close to zero. For what concerns the fluvial layer; n, θs and α show a high sensitivity which is zero in the rest of the simulation period. As expected the saturated permeability of the aquifer layer is the only sensitive hydraulic parameter. In the foundation layer, a significant sensitivity is registered for all the hydraulic parameters during the whole investigated period, expecially the n parameter has a relevant peak of sensitivity in correspondence of flood C. For what concerns the parameters of the wetting curve, their sensitivity is much lower compared to the one of the drying curve and peaks are observed for the fluvial and embankment layers in correspondence of the floods C and G, while for the foundation layer the sensitivity is more homogeneous over the whole simulation period (see Figure 15). Inverse analysis gives in output hydraulic parameters in some cases arranged in a very narrow range around the base values (initial guess of the parameters). It is interesting at this stage to understand if it is significant to consider changes in the third-fifth digit after the comma for parameters θr, θs, n, α and in the fifth-eleventh digit for the saturated permeability in the different layers. To investigate this aspect, a computation of the differences in terms of PH and wc has been performed and the sensitivity coefficient proposed by Saraiva et al (2017) and Gomes et al (2017) has been used between the simulation with the base parameter and the simulation with the changed parameter for every hydraulic parameter and for every layer. Parameters are changed using an increasing number of digits after the comma. It has been observed that the saturated permeability is the parameter to whom variations the model output is more sensible (up to the eleventh digit after the comma) for all the investigated layers, while α and θr have a low (or very low) sensitivity. We understand that when dealing with the saturated permeability K, the saturated water content θs and the shape factor n, it is always suggested to consider even the minimal variation with respect to the initial parameter value.

In the following chapter, the statistical comparison between direct simulations using four different hydraulic models (VGM, HVGM, VGM-2cm, HVGM-2cm) has been presented. Qualitative and quantitative methods have been used for this purpose and pros and cons of each technique have been highlighted.

Afterwards, the analysis has been concentrated expecially on the VGM, which is the simpler model among the adopted ones and the one that requires the lower number of parameters. The quantitative analysis of the performance has been carried out on the single observation point to investigate locally the model performance (in the single layer), globally using the whole set of observation points, and subdivided on the single peak and off-peak periods in order to investigate the model performance in certain periods of the simulation more interesting for our purposes. Moreover the performance of the VGM has been investigated using different datasets of observation point (wc, PH and wc+PH). The last part of chapter 6 focuses on the different typologies of measurements errors that could influence the dataset of observation points (measurement error, sensor accuracy, instrument calibration etc).

6 STATISTICAL COMPARISON BETWEEN DIRECT SIMULATIONS

6_1 STATISTICAL COMPARISON OF THE PERFORMANCE OF DIFFERENT HYDRAULIC MODELS

In the present chapter a comparison between direct simulations of the 2018 period (12th November, 2017 to 23rd June 2018) using four different hydraulic models is presented. The chosen hydraulic models are the following: the van Genuchten model (VGM), the van Genuchten model considering the hysteretic behaviour (HVGM), the van Genuchten model imposing an air-entry value of -2cm (VGM-2cm); the van Genuchten model considering the hysteretic behaviour and an air-entry value of -2cm (HVGM-2cm). The four different hydraulic models available in Hydrus library have been presented extensively in chapters §2_3; §2_4 and §2_5 of the present thesis. The simulations differ from one another only for the hydraulic model, no other changes are introduced for example in the boundary conditions, initial conditions, discretization of the domain etc..

In Table 9 the initial parameters datasets used for the four chosen hydraulic models are reported.

HVGM HVGM-2cm	**θr**	**θs**	**α (1/m)**	**n**	**K (m/min)**	**l**	**θm**	**θsw**	**αw**	**Ksw (m/min)**
Fluvial layer	0,0005	0,319	0,251	1,284	0,0000276	0,5	0,3191	0,287	0,50	0,0000276
Foundation layer	0,001	0,424	0,118	1,142	0,000000198	0,5	0,4241	0,382	0,24	0,000000198
Agrarian layer	0,00005	0,366	0,161	1,307	0,000027	0,5	0,3662	0,330	0,32	0,000027
Aquifer layer	0,01	0,43	0,200	1,200	0,00009	0,5	0,4301	0,387	0,40	0,00009
Subsoil layer	0,00	0,43	0,200	1,200	0,000000078	0,5	0,4301	0,387	0,40	0,000000078
Embankment layer	0,004	0,397	0,850	1,304	0,00009	0,5	0,3975	0,357	1,70	0,00009
VGM VGM-2cm	**θr**	**θs**	**α (1/m)**	**n**	**K (m/min)**	**l**				
Fluvial layer	0,0005	0,319	0,251	1,284	0,0000276					
Foundation layer	0,001	0,424	0,118	1,142	0,000000198					
Agrarian layer	0,00005	0,366	0,161	1,307	0,000027					
Aquifer layer	0,01	0,43	0,200	1,200	0,00009					
Subsoil layer	0,00	0,43	0,200	1,200	0,000000078					
Embankment	0,004	0,397	0,850	1,304	0,00009					

Table 9: Set of parameters used in the four chosen hydraulic models (VGM, HVGM, VGM-2cm, HVGM-2cm)

For each layer, observation points are chosen where reliable sensors monitoring the water content (wc) or pressure head (PH) are available in situ. The chosen sensors are sufficient to give to the modeller enough information on the pwp distribution map in time of the investigated bank section. In chapter §6_2, to evaluate the performance of direct simulations, only a small part (7 sensors) of the available installed sensors has been used. A larger number of sensors will be used in chapter §6_2 to investigate in detail the performance of the direct simulation that uses the VGM, while the whole set of sensors will be used in the calibration phase to perform the inverse analysis and to evaluate the performance of each indirect simulation.

In particular, the seven observation points chosen for the present chapter are the following:

- For the embankment layer: MPS6-MPC1-4,6 m, MPS6-SPC1-7m (pore pressure measurements), GS3-MPC1-4,5m and GS3-SPC2-7,1 m (water content measurements).
- For the foundation layer: T8-TB1-4,9m (pore pressure measurements). No sensors for the water content monitoring are available in the foundation layer.
- For the fluvial layer: GS3-MPB2-2,2m (water content measurements). A sensor monitoring the pore pressure has been installed in the fluvial layer (MPS6_SPB1_1,2m) but the author prefers to not use these data in this phase because a strong influence of the evapo-transpirative processes is detected at this depth and, for this reason, the use of different hydraulic models affects only marginally this observation point.
- For the aquifer layer: PZ1-17 m (pore pressure measurements). No sensor for water content monitoring is available in the aquifer layer.

In Figure 17, seven graphs, one for each selected observation point, are presented. In each graph, the observed data are compared to the simulated data for each of the four hydraulic models. As it is possible to observe:

- The models reproduce satisfactorily MPS6-SPC1 7,0m (embankment), T8-TB1-4,9m (foundation), PZ1-17m (aquifer) following the peaks and the trend of the observation measurements. A very low bias is observed for these observation points. MPS6-SPC1-7,0m despite an initial condition far from the one observed in situ, in time showed a reduction of the influence of the initial conditions and a recovery of the experimental trend.
- The models reproduce relatively well GS3-SPC2-7,1m and GS3-MPB2-2,2m dataset and the bias between observed and simulated data is due to a not perfect match with the initial observed water content while the overall trend in time is followed. The influence of the initial conditions does not seem to reduce in time.
- For what concerns GS3-MPC1-4,5m (embankment) and MPS6-MPC1-4,6m (embankment), the models seem to be responsive to the seventh flood of the 2018 period (period 158200-193000min, from 1/03/2018-26/03/2018, duration~25 days), showing a very steep increase both in pore pressure and in water content at the same depth (4,5 m) that is not observed in the in situ data. The embankment seems less responsive to external hydraulic stimuli with respect to the simulated behaviour of the model.
- All the considered hydraulic models show almost the same simulation trends for the investigated observation points (same peaks, same curve decay etc). The differences between the simulation

trends of models VGM, HVGM, VGM-2cm are extremely reduced while HVGM-2cm shows the greatest differences.

In Figure 18, observed data (x-axis) versus simulated data (y-axis) represent the scatter of the data points around the line of perfect agreement (1:1). This is another interesting way to look at the same information reported in Figure 17. As could be observed T8-TB1-4,9m, PZ1-17m, MPS6-SPC1-7m present an acceptable scattering from the best fit (45° black line) while MPS6-MPC1-4,6m and GS3-MPB2-2,2m show a constant underestimation of the data and GS3-MPC1-4,5m an overestimation of the observed data.

What has just been exposed is the standard qualitative procedure adopted to compare graphically different simulations or verify a trial and error calibration procedure. But, as could be seen, from a graphical comparison, it is extremely difficult to quantify the performance of a group of simulations (that use for example different hydraulic models) or the performance of a model with respect to different in situ observations.

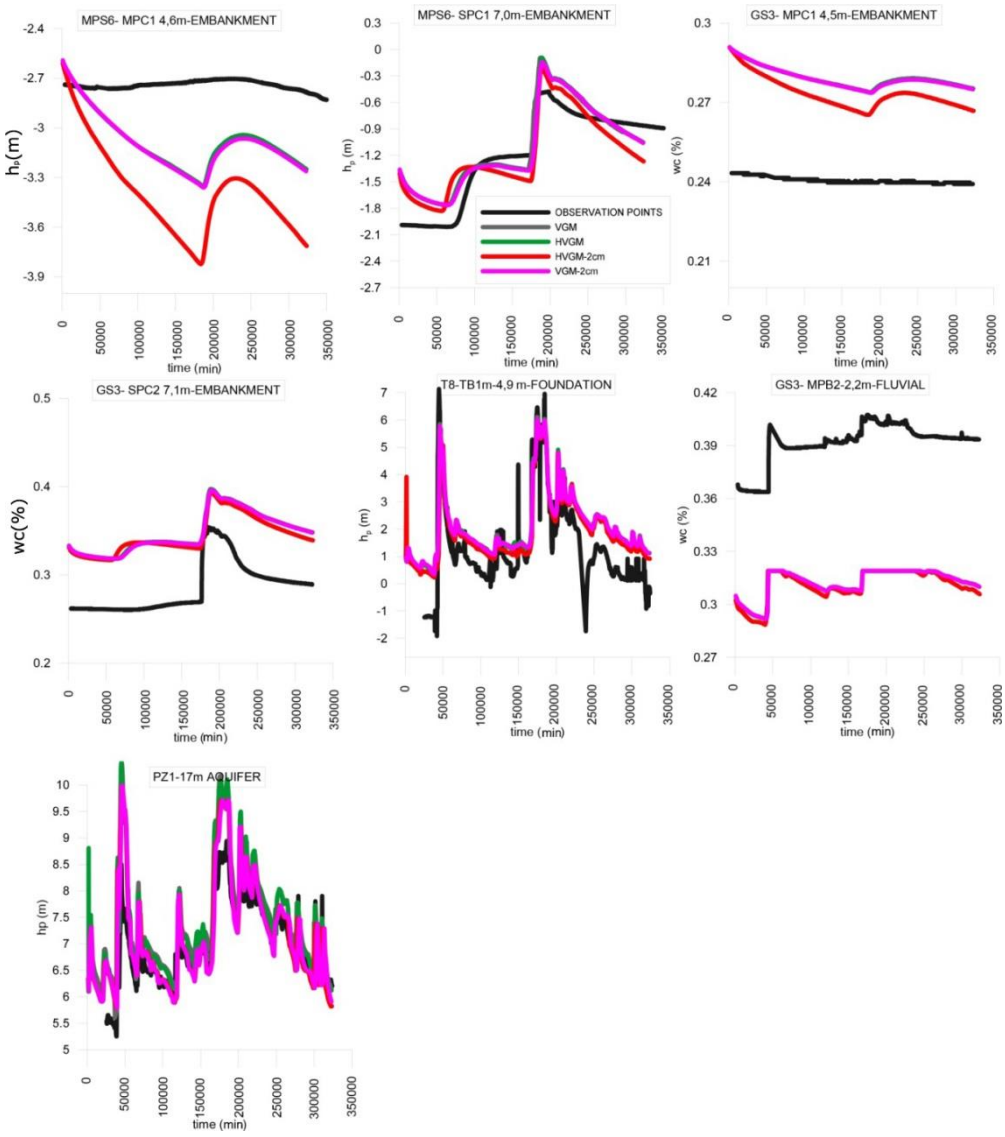


Figure 17: Comparison between the observed dataset (black points) and the simulated dataset using four different hydraulic models (VGM, HVGM, VGM-2cm, HVGM-2cm) in the considered 7 observation points (MPS6_MPC1_4,6m; MPS6_SPC1_7m; GS3_MPC1_4,5m; GS3_SPC2_7,1m; T8_TB1_4,9m; GS3_MPB2_2,2m, PZ1_17m).

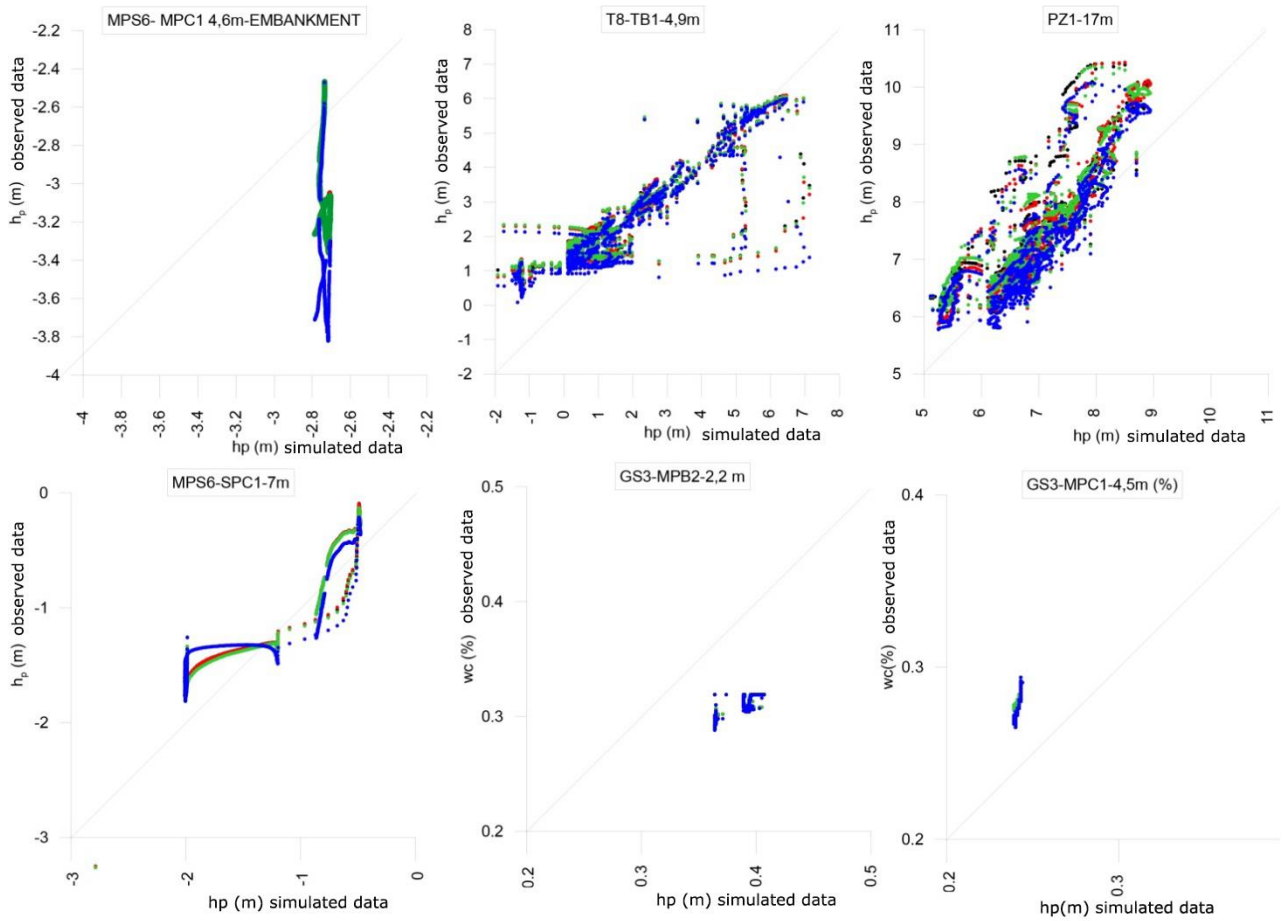


Figure 18: Graphs representing the scattering of the data points observed-simulated around the best fit line (1:1) of the four considered hydraulic models

Figure 17 and Figure 18 show reduced differences in terms of pressure head and water content between the different hydraulic models. At this stage, an effort will be made to understand the entity of those differences in order to understand if an increase in the complexity of the model passing from a VGM with 6 parameters to an hysteretic model with 10 hydraulic parameters could be meaningful and desirable for our purposes.

In Table 10 the maximum deviations (in the positive and negative range) between the VGM (taken as the reference) and the other considered hydraulic models (HVGM, VGM-2cm; HVGM-2cm) are presented. The hydraulic model HVGM-2cm shows the highest differences for almost all the selected points, while HVGM and VGM-2cm show relevant changes only in the observation points PZ1-17 m (subsoil layer) and T8_TB1_4,9m (fluvial layer/berm). HVGM seems to respond quite differently in the berm (fluvial layer) with respect to the VGM while the response in the embankment layer is almost the same. Table 10 is able to underline the entity of the maximum differences between the different hydraulic models but it does not give information in which temporal instants of the simulation those differences are recorded.

	HVGM - VGM	VGM-2cm - VGM	HVGM-2cm - VGM
PZ1_17m (m)	0,428/-0,940	0,149/-0,124	0,580/-0,749
T8_TB1_4,9m (m)	0,159/-0,583	0,044/-0,085	0,140/-3,355
MPS6_MPC1_4,6 m (m)	0,004/-0,002	0,002/-0,03	-0,004/-0,482
MPS6_SPC1_7m (m)	0,023/-0,005	0,007/-0,04	0,252/-0,210
GS3_MPB2_2,2 m (-)	0,002/-0,001	0,001/-0,001	0,020/-0,005
GS3_MPC1_4,5m (-)	0,001/-0,001	0,001/-0,001	0,000/-0,009
GS3_SPC2_7,1 m	0,001/-0,001	0,001/-0,002	0,012/-0,009

Table 10: Maximum deviations registered during the investigated period (2018) between the VGM (taken as reference) and the considered hydraulic models (HVGM, VGM-2cm, HVGM-2cm). In red the values that show the highest differences between the VGM and the other hydraulic models.

Adopting the hysteretic models (HVGM or HVGM-2cm) instead of the VGM or VGM-2cm could be considered relevant (despite the higher calculation time, the convergence problems and the higher number of parameters to be properly calibrated) if relevant changes are detected during the major flood events (long-lasting high river water level).

It is interesting, at this stage, to understand when the major differences are observed between the hysteretic (HVGM) and non-hysteretic model (VGM) during the period under investigation and give a possible explanation of it. The same differences could be found for the hysteretic model (HVGM-2cm) compared to the non-hysteretic model (VGM-2cm) but results are not shown here for reason of brevity.

Contrary to what could be expected, the inclusion of the hysteretic effects provide just limited changes in the embankment behaviour (as could be observed in Table 10) while greater changes are detected for the fluvial layer (berm). The explanation could be found in Liu et al (2015) that tested numerically an embankment under simplified sinusoidal variations of water level adopting a non-hysteretic model and a hysteretic model. It has been noted that the duration of a flood event is a factor affecting greatly the hysteretic behaviour of a river embankment. In fact if the period of the flood is small compared to the permeability of the embankment, volumetric water content and pore pressure changes have little time to propagate and their variations are limited to very small ranges. If we compare the water level fluctuation (in m a.s.l) and the deviation in the response of the HVGM compared to VGM in the observation points (see Figure 19), it is clear that the greatest variations are detected in correspondence of the third flood C (period 35000-60000 min, 6/12/17 to 23/12/2017, duration ~17 days), the fifth flood E (period 116400- 139650min, 31/01/18 to 16/02/18, duration ~16 days), seventh flood G (period 158200-193000 min, 22/02/18 to 26/03/18, duration~24 days) and eighth flood H (period 197970-238140, 29/03/18 to 26/04/18, duration ~28 days).

A part from the subsoil that is a very deep layer (PZ1_17m), the most relevant differences are observed for the fluvial layer (berm) as already highlighted in Table 10 because the hydraulic stimuli affect earlier this area and here they are more persistent in time (with respect to the embankment layer). Variations in pore pressure due to hysteretic behaviour have to be considered when computed for stability analysis because the factor of safety (FOS) is always overestimated in non-hysteretic models leading to potentially dangerous situations (Liu et al, 2015).

In Table 11 a quantification of the performance of the HVGM compared to the VGM is presented by means of a set of metrics during the flood periods highlighted in Figure 19: floods C; E and F, G, H. Focus the attention from the whole simulation period to discrete temporal intervals could be extremely useful in order to investigate the model performance in interesting time periods for the purpose of the analysis.

Accuracy metrics (MBE, MAE, RMSE, R^2), efficiency metrics (NSE, IA, NSE_j, IA_j, NSE_{rel}, IA_{rel}, IA') and statistical significance criteria (p-value) are chosen as indices. The chosen 7 observation points are the same used in Table 10 and Figure 19 (MPS6_MPC1_4,6m; MPS6_SPC1_7m, GS3_MPC1_4,5m; GS3_SPC2_7,1m; T8_TB1_4,9m; GS3_MPB2_2,2m; PZ1_17m). As it could be easily read comparing the values of the indices to the best fit (2nd column), the HVGM shows a better performance compared to VGM for the flood periods C, E and F, H. The differences between the indices of the two models are reduced as could be read also in Figure 19. The general performance of both the investigated models is satisfactory.

A set of metrics/indices could be used also to give a quantitative comparison of the differences in the observation points of the berm, highlighted qualitatively in Figure 19, between the VGM and HVGM (see Table 12). In some cases it is interesting to investigate the behaviour of the simulation locally in particular observation points instead of considering the behaviour of the simulation on the whole group of observation points. This allows to have a more refined investigation of the model performance in particular zones of the domain whose problems could not be highlighted facing the simulation globally instead of locally. The chosen observation points are localized in the berm: T8_TB1_4,9m (PH) and GS3_MPB2_2,2m (wc). Comparing the values of the metrics with the best fit (2nd column) it is clear that the HVGM is able to reproduce better the pwp and the wc distributions in the fluvial layer (berm), as we could expect. These differences, as highlighted multiple times, are small but significant because detected in the most relevant time periods of the analysis (major flood peaks).

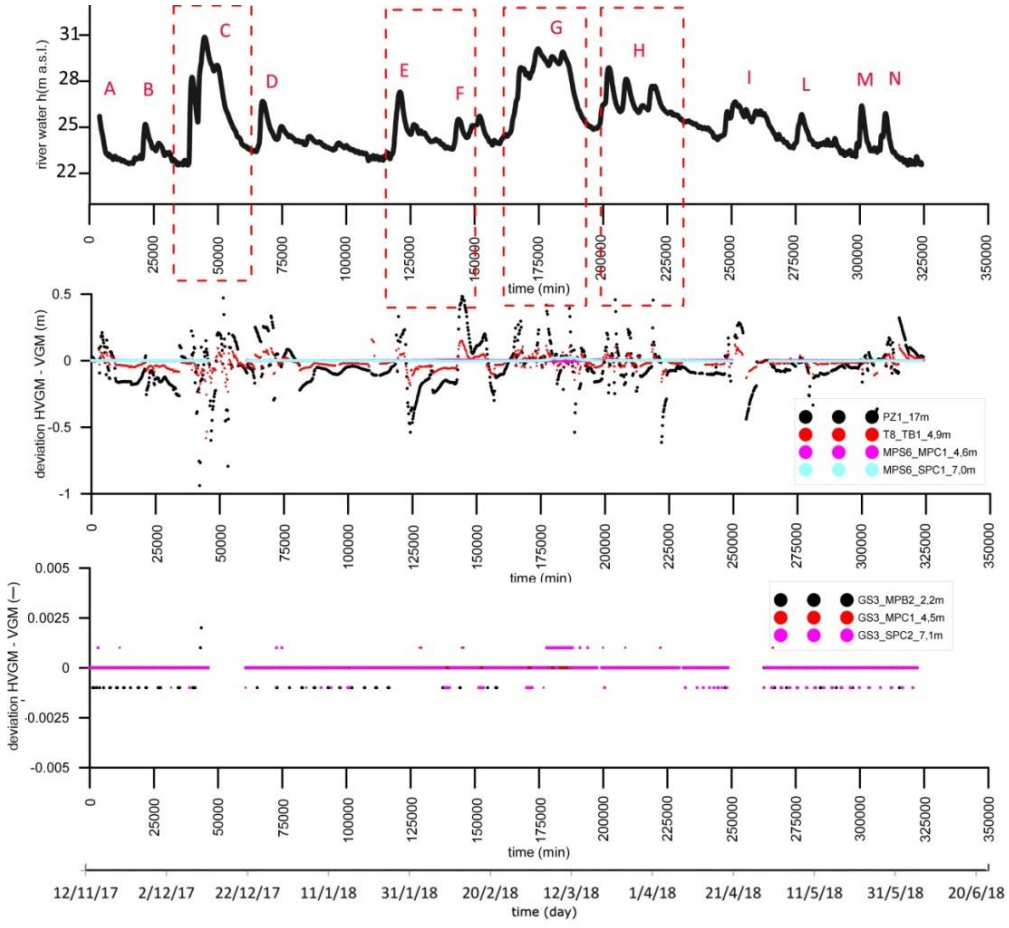


Figure 19: In the top chart the river water height (m a.s.l) vs time with indication of the flood events using letters from A to N. In the chart in the centre the differences (in terms of pressure head (m)) between the VGM and HVGM in the observation points MPS6-MPC1-4,6m; T8-TB1-4,9m, MPS6-SPC1-7m and PZ1-17m. In the bottom chart, the differences (in terms of water content)

between the VGM and HVGM in the observation points MPB2_2,2m; MPC1_4,5m; SPC2_7,1m. A red dashed line highlights the temporal periods in which greater differences are found between the two considered hydraulic models.

		VGM				HVGM			
	BEST FIT	C	E-F	G	H	C	E-F	G	H
MBE	0	-0,749	-0,125	-0,151	-0,279	-0,691	-0,108	-0,159	-0,269
MAE	0	1,216	0,434	0,572	0,453	1,178	0,405	0,576	0,441
RMSE	0	1,611	0,546	0,745	0,510	1,587	0,525	0,758	0,499
R^2	1	0,898	0,985	0,987	0,994	0,896	0,986	0,986	0,994
NSE	1	0,835	0,979	0,971	0,985	0,839	0,980	0,970	0,986
IA	1	0,963	0,995	0,994	0,997	0,963	0,995	0,993	0,997
NSErel	1	0,666	0,850	0,960	0,977	0,677	0,854	0,960	0,977
IArel	1	0,924	0,965	0,991	0,995	0,927	0,966	0,991	0,995
IA’	1	0,827	0,934	0,929	0,937	0,833	0,938	0,928	0,939
NSEj	1	0,654	0,867	0,858	0,875	0,666	0,876	0,857	0,878
IAj	1	0,836	0,934	0,932	0,939	0,841	0,938	0,931	0,941
p_value	>0,05	0,070	0,605	0,612	0,280	0,093	0,655	0,596	0,298

Table 11 Set of metrics used to compare the behaviour of the VGM and HVGM during the major flood periods (flood C, E-F, G, H). The chosen observation points are the same used in Figure 19.

		T8_TB1_4,9m (PH)		GS3_MPB2_2,2m (wc)	
	BEST FIT	VGM	HVGM	VGM	HVGM
MBE	0	-0,8709	-0,8540	0,0788	0,0788
MAE	0	1,0072	0,9915	0,0788	0,0788
RMSE	0	1,2171	1,2079	0,0790	0,0791
R^2	1	0,8434	0,8690	0,7741	0,7768
NSE	1	0,5783	0,5847	-41,240	-41,296
IA	1	0,8633	0,8652	0,2028	0,2028
NSErel	1	-3,156	-3,055	-41,013	-41,073
IArel	1	-0,3469	-0,316	0,2071	0,2071
IA’	1	0,6398	0,6455	-3,6151	-3,6184
NSEj	1	0,2797	0,2909	-8,2301	-8,2368
IAj	1	0,5855	0,5906	0,0978	0,0977

Table 12 Set of metrics used to compare the behaviour of the VGM and HVGM in the observation points located in the berm (T8_TB1_4,9m (PH); GS3_MPB2_2,2m (wc)).

		DATASET: PH				DATASET :wc			
	BEST FIT	VGM	VGM-2cm	HVGM-2cm	HVGM	VGM	VGM-2cm	HVGM-2cm	HVGM
MBE	0	-0,129	-0,822	-0,822	-0,124	-0,005	-0,005	0,0008	-0,005
MAE	0	0,523	1,721	1,721	0,518	0,048	0,048	0,048	0,048
RMSE	0	0,738	2,035	2,035	0,733	0,055	0,055	0,054	0,055
R^2	1	0,895	0,606	0,610	0,896	0,378	0,379	0,396	0,377
NSE	1	0,844	-0,181	-0,181	0,846	0,250	0,252	0,258	0,249
IA	1	0,966	0,399	0,399	0,966	0,676	0,677	0,702	0,676
NSErel	1	0,323	0,774	0,774	0,339	0,357	0,359	0,373	0,356
IArel	1	0,852	0,885	0,885	0,855	0,722	0,723	0,748	0,722
IA'	1	0,811	0,377	0,377	0,813	0,569	0,569	0,564	0,568
NSEj	1	0,622	-0,245	-0,245	0,625	0,138	0,139	0,128	0,137
IAj	1	0,834	0,252	0,252	0,836	0,472	0,473	0,486	0,471
AIC/ AICmin	1	1,01	1,92	1,92	1,00	1,159	1,160	1,163	1,158
BIC/ BICmin	1	1,01	1,92	1,92	1,00	1,157	1,159	1,162	1,157
HQC/ HQCmin	1	1,01	1,92	1,92	1,00	1,158	1,160	1,163	1,158
p_value	>0,05	0,25	1,2e-172	1,2e-172	0,40	1,39E-5	2,5e-05	0,502	1,4E-5

Table 13: Set of metrics applied to compare the performance of the different hydraulic models (VGM, HVGM. VGM-2cm, HVGM-2cm) and the different datasets (PH, wc) applied to the simulation of the 2018 period.

In Table 13 a set of metrics has been used in order to give an objective evaluation of models performance for a direct comparison. Metrics have been calculated for each different hydraulic model (VGM, HVGM, HVGM-2cm, VGM-2cm) and using different datasets of observation points: a PH dataset (PZ1_17m; TB1_4,9m; MPC1_4,6m; SPC1_7m, MPC3_6,2m) and a wc dataset (MPB2_2,2m; MPC1_4,5m; SPC2_7,1m, MPC1_2,4m). The same observation points are used for Figure 17, Figure 18 and Figure 19 of the present chapter.

For what concerns the error metrics (MBE, MAE, RMSE), the greater bias belongs to the VGM-2cm and HVGM-2cm for a PH dataset while the water content dataset shows an almost homogeneous behaviour in all the four considered models. The comparison between water content dataset and pressure head dataset reveals that the accuracy metrics (MBE, MAE, RMSE) show smaller error values when calculated on a water content dataset compared to the values obtained for a pressure head dataset. This is easily explained by the fact that MBE, MAE, RMSE have the same units of the input data (measured and simulated data) and water content and pressure head differ by one order of magnitude. For this reason, it is not possible a direct comparison between datasets with different unit of measure using MBE, MAE, RMSE without a prior proper normalization (as performed for the calibration programme in chapter §7 according to the normalization proposed in chapter §2_7_2_1). As could be observed MBE has the lowest dispersion due to the logical fact that the residuals keep their signs (and could be therefore cancel out) while MAE and RMSE are always sum of positive errors.

RMSE of the PH dataset shows a larger error compared to MAE because errors are squared in RMSE before being averaged. RMSE varies with the square root of the number of errors (n), with the variability of the error magnitudes within the dataset and with the magnitude of the averaged errors (as MAE). This metric gives an higher weight to large errors and for this reason, MAE could be considered a more natural measure of the average error in order to avoid unrealistic model prediction assessments (Phogat et al, 2016). Despite the higher error indicated by RMSE metric for the PH dataset, the low value of MBE and MAE conceals the simulation inaccuracies. From all this comes the necessity to consider more than one

accuracy metric at a time. One of the main limits of MAE and RMSE is the lack of indication of the error direction, which is given on the contrary by MBE. This limitation could be ignored if the focus of the analysis is on the magnitude of the error and not on its direction. From the values of the accuracy errors in Table 13, it is possible to state that all the considered models show a good performance in the investigated 2018 period.

The coefficient of determination R^2 communicates analogous information of the accuracy metrics, with a value close to 0,38 for the water content dataset (performance non acceptable) and close to 0,89 for the PH dataset. Despite the good behaviour shown in the present case study, it is strongly suggested to not use R^2 as the only indicator for simulation evaluation because it only quantifies the dispersion between the dataset values but it has been demonstrated that if the model systematically over or under predict all the times, the result of this metric could be very close to 1 even if all predictions are incorrect (Krause et al, 2005). Moreover R^2 is over sensitive to extreme values and insensitive to additive and proportional differences between observed-simulated data (Legates and McCabe, 1999).

The efficiency criteria show a very good general fit between observed and simulated datasets with error values extremely low (efficiency metrics close to the best fit). The larger error shown by NSE_{rel} in Table 13 for both the types of dataset (with respects to the other efficiency metrics) reflects the wide differences between measured and simulated data in certain times of the simulation, but the use of relative deviations are able to reduce the influence of the absolute differences between measured-simulated. This is due to the sensitivity of NSE_{rel} to large variations in values for which a higher weighting is assumed, while small divergences are neglected. In other words, larger errors are emphasized while small errors neglected. The negative value of NSE and NSE_j obtained for the pore pressure dataset shows a high under prediction of the models that is not confirmed from the other efficiency metrics values. The behaviour of the remaining efficiency criteria (IA, IA' IAj) is homogeneous: lower efficiency for the water content dataset with small differences between the hydraulic models, greater value of efficiency for the pressure head dataset in particular for the VGM and HVGM.

The model selection criteria AIC/AICmin, BIC/BICmin, HQIC/HQICmin are particularly useful to perform a probabilistic selection of the best model among many. These metrics are able to take into consideration the model performance on the training dataset and the complexity of the model (number of parameters). The best fit for the model criteria is 1 and the best model is the one that get closer to the best fit. For both the datasets (wc and PH), the VGM and HVGM are the models with the best performance according to the model selection criteria.

The level of statistical significance p-value could be used to investigate the significance of two independent variables or datasets. The assumed null hypothesis is that the two variables/databases are dependent and belong to the same family (no statistical difference). If the p-value is below the significance level 0,05, the null hypothesis is rejected. Looking at the p-values in Table 13, the HVGM-2cm is the only model that does not reject the null hypothesis for the wc dataset while VGM and HVGM for the PH dataset.

The combined use of all these metrics together, each of them with its pros and cons, assures a good objective assessment of the closeness of the simulated behaviour to the observed datasets and it allows also comparing quantitatively different models. From the comparison between hydraulic models (VGM. HVGM. VGM-2cm, HVGM-2cm) under the same observation dataset (PH or wc) through the same set of metrics in Table 13, it is possible to observe that the VGM and HVGM show the best performance for a PH

dataset while VGM-2cm and HVGM-2cm for a wc dataset (better agreement between observed-simulated data).

In the following chapter §6_2, the attention will be focused on the VGM because it is a simple but reliable model implemented in Hydrus that requires a lower number of parameters compared to the models considering the soil hysteretic behaviour. The statistical analysis performed in chapter §6_2 and proposed for the VGM by way of example, could be applied in the same way for the remaining hydraulic models. While in subchapter §6_1 a restricted number of observation points (7) has been used in order to facilitate the comparison between the different hydraulic models, in subchapter §6_1 an higher number of observation points has been used in order to perform an in-depth analysis of the performance of the VGM.

6_2 STATISTICAL ANALYSIS OF THE PERFORMANCE OF THE VGM

The analysis of the simulation performance of the VGM presented in the previous chapter §6_1, Table 13 has been performed considering a restricted number of sensors of the same typology (PH or wc) together. This allows having a general overview on the performance of the model in the domain of interest. When comparing different models, as done in Table 13, it is recommended to use a restricted number of sensors on whose readings the modeller has high confidence and, in case, to choose a restricted set of metrics tested in advance on similar datasets. All this to avoid confusion and disorientation facing the enormous amount of obtained data. Once the modeller chooses the best simulation among many, a more in-depth investigation could be performed on the selected one.

The chosen observation points are the following:

-wc dataset embankment layer: GS3_SPC2_7,1m; GS3_MPC1_2,4m; GS3_MPC1_4,5m; GS3_MPC3_6,4m

-PH dataset embankment layer: MPS6_MPC1_4,6m; GS3_MPC3_6,2m; MPS6_SPC1_7m, T8_TC2_8m

-PH dataset foundation layer: T8_TB1_4,9m; MPS6_MPB2_2,7m; PZ2_10m

-wc dataset fluvial layer: GS3_MPB2_2,2m.

-PH dataset aquifer layer: PZ1_17m.

In Table 14 the overall performance of the VGM on the whole investigated time period (2018 year) has been presented using the new set of observation points. As could be observed, the VGM has an excellent performance in the simulation of the pwp distribution in time, with values of the metrics in all cases very close to the best fit. This suggests that the initial dataset of the hydraulic parameters, obtained as average values of the laboratory parameters, is well representative of the in situ soil properties. We can expect that the calibration process is not going to enhance greatly the performance of the simulation in terms of pressure head, while huge room for improvement is possible for the simulation in terms of wc. In fact the chosen metrics suggest a poor performance of the simulation in terms of water content distribution in the investigated section, with values of the indices in most of the cases below acceptability.

In some cases it is interesting to focus the attention on the performance of the single sensors in order to individuate possible localized problems such as inaccurate description of the phenomena in certain layers or zones of the domain. In Table 15 a set of metrics is applied to single observation points located in the embankment layer, half of them are sensors monitoring the pwp, the other half sensors monitoring the wc.

For what concerns the PH sensors, the simulation shows a good performance for almost all the considered observation points, in agreement with what has been found in Table 14. An exception is the observation point MPS6-MPC1-4,6m that shows values of the efficiency metrics less than zero. It is commonly known that values of efficiency metrics below 0 are undesirable since they imply that the mean of the observed system output is a better predictor then the model itself. In Figure 17, it is possible to observe that the simulated trend of MPS6-MPC1-4,6m shows a higher responsivity to hydraulic stimuli with respect to the observed trend. The simulation of the behaviour of the observation points (type wc) is poor as highlighted previously in Table 15. Here as well, the values of the efficiency metrics are close to 0 or below zero, while the values of the accuracy metrics are close to the best fit because in the same unit of measurement of the wc dataset (values between 0 and 1). It is interesting to stress that the coefficient of determination R^2 shows values close to the best fit (1) for the wc observation points. Several reason could lead to an overestimation of the coefficient of determination as a problem of model overfitting (model is too complex to be described by the dataset of observations) (Babyak, 2004). It is important to remember that R^2 is always a biased estimator (it is systematically too high or low) and for this reason, it is recommended the use of this index together with others.

VGM					
	BEST FIT	PH dataset		BEST FIT	wc dataset
MBE	0	-0,279	MBE	0	-0,013
MAE	0	0,528	MAE	0	0,047
RMSE	0	0,770	RMSE	0	0,053
R^2	1	0,954	R^2	1	0,244
NSE	1	0,939	NSE	1	0,150
IA	1	0,986	IA	1	0,653
NSE_{rel}	1	0,782	NSE_{rel}	1	0,250
IA_{rel}	1	0,949	IA_{rel}	1	0,693
IA'	1	0,893	IA'	1	0,503
NSE_j	1	0,785	NSE_j	1	0,007
IA_j	1	0,894	IA_j	1	0,430

Table 14: Set of metrics applied to the direct simulation that uses the VGM. The performance of the simulation is investigated in terms of pressure head and water content.

VGM	BEST FIT	MPS6 MPC1 4,6m (Emb_PH)	MPS6 MPC3 6,2 m (Emb_PH)	MPS6 SPC1 7,0 m (Emb_PH)	T8 TC2 8,0 m (Emb_PH)	GS3 MPC1 2,4 m (Emb_wc)	GS3 MPC1 4,5 m (Emb_wc)	GS3 MPC3 6,4 m (Emb_wc)	GS3 SPC2 7,1 m (Emb_wc)
MBE	0	0,353	0,248	-0,107	-0,448	0,001	-0,039	-0,044	-0,063
MAE	0	0,363	0,313	0,194	0,448	0,012	0,039	0,044	0,063
RMSE	0	0,396	0,358	0,234	0,507	0,014	0,039	0,044	0,064
R^2	1	0,060	0,650	0,853	0,735	0,878	0,804	0,810	0,867
NSE	1	-358,52	0,305	0,792	-0,475	-136,92	-1030	-4,742	-3,908
IA	1	0,066	0,817	0,941	0,740	0,256	0,056	0,509	0,529
NSE_{rel}	1	-360,03	0,086	0,682	-3,015	-136,06	-1029	-4,998	-4,209
IA_{rel}	1	0,062	0,759	0,909	0,291	0,2606	0,057	0,487	0,500
IA'	1	-9,433	0,579	0,771	0,335	-5,270	-18,587	-0,382	-0,305
NSE_j	1	-19,866	0,159	0,543	-0,329	-11,53	-38,17	-1,764	-1,609
IA_j	1	0,044	0,602	0,761	0,438	0,134	0,025	0,266	0,277

Table 15: Set of metrics applied to the VGM simulation to investigate the performance of the single sensor located in the embankment layer.

In the previous analysis (chapter §6_1), the attention has been focused mainly on the performance of the direct simulations during the whole 2018 period, but, at this stage, it is interesting to study the behaviour

of the direct simulation (VGM) over more restricted periods f.e. the major flood events (for river water height and persistence in time). It must indeed be kept in mind that the final goal is the elaboration of a reliable pwp distribution map during the most critical time instants of the simulation, for future possible stability analysis.

For this reason, instead of considering the performance of the investigated observation points on the whole 2018 period, the time axis is divided in 12 periods of ''flood peak'' and 12 periods of ''off-peak''. 25 m a.s.l is considered the reference river water level to individuate the major flood periods. The individuated ''peak periods'' are indicated with capital letters from A to N, while the ''off-peak periods'' with letters from 'A' to O' (see Figure 20). In Table 16 the initial and final temporal instants of each ''peak-periods'' and ''off-peak periods'' are presented in detail (the temporal scale in minutes is based on the investigated 2018 period from the 12th November 2017 which is the zero instant, to the 23rd June 2018, the last temporal instant). The chosen observation points are located in the berm and in the embankment. Half of the observation points are associated to pore pressure readings and the other half to water content readings. Sensors installed in the foundation and in the subsoil layers have not been considered in order to focus the attention on the most interesting layers for the purposes of the present work. The chosen observation points are the following:

- For the fluvial (berm) and foundation layers: MPB2_2,2m (wc sensor); T8_TB1_4,9m (PH sensor) have been chosen.
- For the embankment layer: MPS6_MPC1_4,6m (PH sensor); MPS6_MPC3_6,2m (PH sensor); MPS6_SPC1_7m (PH sensor); SPC2_7,1m (wc sensor); MPC1_2,4m (wc sensor); MPC1_4,5m (wc sensor) have been chosen.

The chosen set of metrics is restricted to two accuracy indices (RMSE, R^2), two efficiency indices (NSE, IA') and a statistical significance criteria (p-value), because, as seen in chapter §4, together they are able to give a good and reliable quantification of a simulation performance. The use of a greater number of metrics is not recommended when dealing with multiple time intervals because the modeller could be lost in the huge quantity of information, not drawing any conclusion.

As could be observed in Table 17, the investigated model is able to reproduce with high accuracy the pressure head dataset both in the ''peak periods'' and in the ''off-peak periods''. The most relevant flooding events (floods C, G, H) are very well simulated by the VGM with a coefficient of determination around 0,85 for flood C and 0,98 for G and H, while the p-values highlight that the simulated and observed datasets belong to the same family and are not statistically different (the majority of the p-values is above the significance level of 5%).

The values of the set of metrics suggest a poorer performance of the model using the wc dataset, with R^2 values below 0,5 (unacceptable performance) and efficiency criteria between -0,1 and 0,6 for both the time periods (peak and off-peak periods). The investigated VGM during the ''off-peak'' periods shows the worst performance, far below the acceptance threshold. All this aspects have to be considered in the calibration phase in order to focus the attention on the most relevant problems detected in the investigated simulations, always keeping in mind the main goals of the simulation.

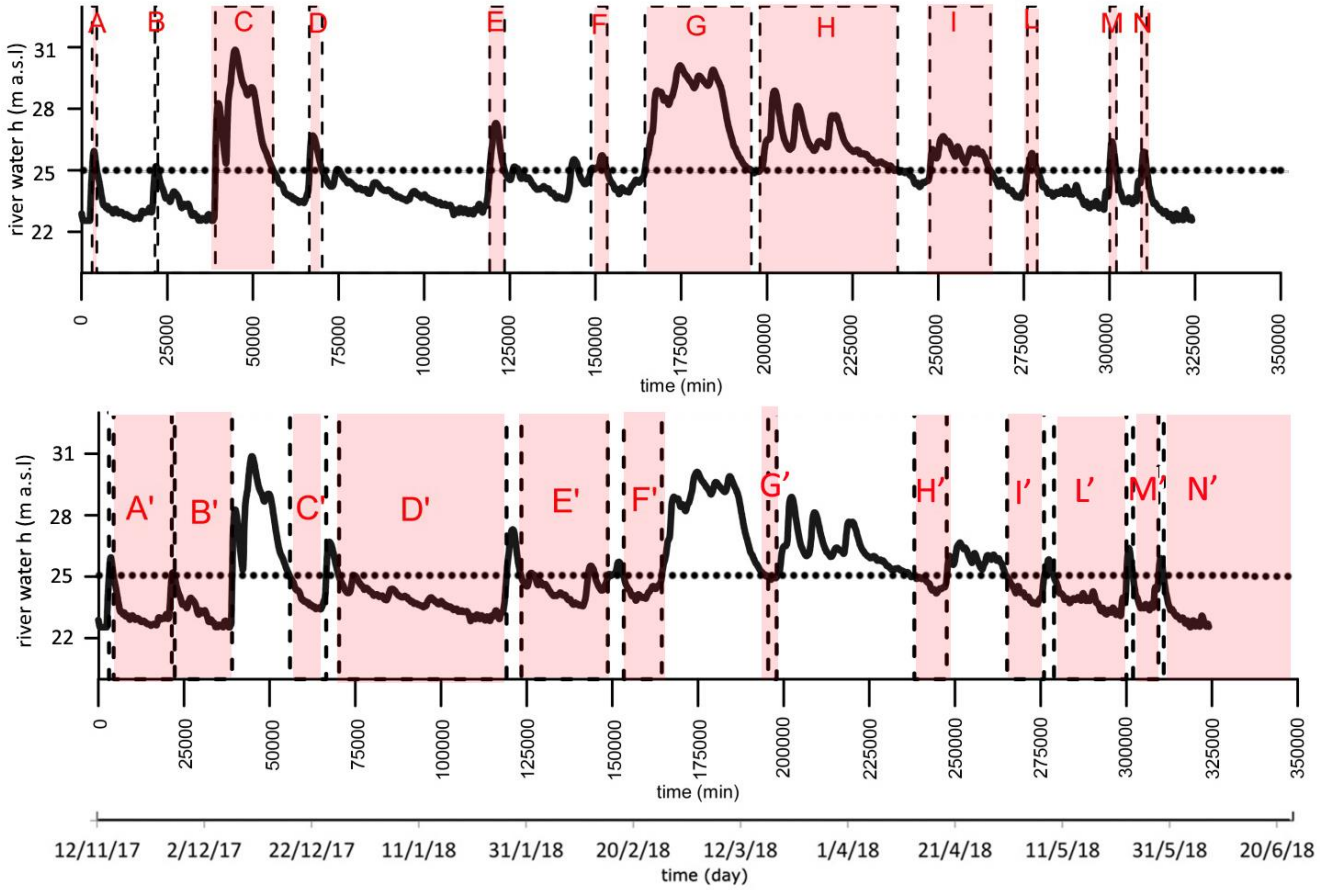


Figure 20: Graphical indication of the considered flood peak periods and off-flood peak periods.

PEAK	From (min)	To (min)	From (day)	To (day)	OFF-PEAK	From (min)	To (min)	From (day)	To (day)
A	3090	4470	14/11/17	15/11/2017	**A'**	4471	21419	15/11/17	26/11/17
B	21420	22260	26/11/17	27/11/17	**B'**	22261	39029	27/11/17	9/12/17
C	39030	55920	9/12/17	20/12/17	**C'**	55921	66479	20/12/17	28/12/17
D	66480	70260	28/12/17	30/12/17	**D'**	70259	119099	30/12/17	02/02/18
E	119100	123450	02/02/18	05/02/18	**E'**	123451	148679	05/02/18	23/02/18
F	148680	153360	23/02/18	26/02/18	**F'**	153361	164399	26/02/18	06/03/18
G	164400	195450	06/03/18	27/03/18	**G'**	195451	197969	27/03/18	29/03/18
H	197970	238140	29/03/18	26/04/18	**H'**	238141	247589	26/04/18	02/05/18
I	247590	265230	02/05/18	15/05/18	**I'**	265231	275999	15/05/18	22/05/18
L	276000	278880	22/05/18	24/05/18	**L'**	278881	300029	24/05/18	08/06/18
M	300030	301980	08/06/18	09/06/18	**M'**	301981	309359	09/06/18	14/06/18
N	309360	310890	14/06/18	15/06/18	**N'**	310891	324040	15/06/18	23/06/18

Table 16: Table with temporal indication of the flood peak periods and off-flood peak periods in terms of minutes (from the zero instant, 12th November 2017) and days.

Pressure Head dataset												
	A	B	C	D	E	F	G	H	I	L	M	N
RMSE	1,458	1,162	1,283	0,682	0,476	0,664	0,569	0,422	0,896	0,761	0,963	0,900
R^2	0,950	0,957	0,856	0,990	0,980	0,885	0,974	0,991	0,986	0,927	0,909	0,919
IA'	-0,109	0,271	0,890	0,837	0,827	0,794	0,920	0,898	0,709	0,648	0,535	0,547
NSE	-6,01	-2,811	0,855	0,826	0,890	0,814	0,965	0,957	0,613	0,515	0,181	0,297
p_value	0,008	0,138	0,810	0,397	0,770	0,362	0,311	0,539	0,045	0,863	0,563	0,000
	A'	B'	C'	D'	E'	F'	G'	H'	I'	L'	M'	N'
RMSE	1,168	1,039	0,552	0,508	0,489	0,439	0,360	0,953	0,668	0,764	0,709	0,843
R^2	0,946	0,954	0,992	0,982	0,968	0,992	0,990	0,814	0,944	0,926	0,935	0,871
IA'	0,137	0,308	0,878	0,810	0,820	0,845	0,909	0,635	0,686	0,632	0,671	0,531
NSE	-3,506	-2,311	0,910	0,853	0,884	0,912	0,962	0,226	0,653	0,518	0,639	0,318
p_value	0,000	0,000	0,240	0,366	0,419	0,255	0,932	0,211	0,838	0,372	0,800	0,456
Water content dataset												
	A	B	C	D	E	F	G	H	I	L	M	N
RMSE	0,054	0,052	0,051	0,051	0,059	0,058	0,053	0,054	0,056	0,055	0,055	0,055
R^2	0,064	0,040	0,174	0,381	0,168	0,198	0,439	0,404	0,306	0,299	0,285	0,286
IA'	0,421	0,426	0,486	0,563	0,547	0,552	0,670	0,619	0,531	0,528	0,522	0,522
NSE	-0,096	-0,085	0,119	0,315	0,134	0,174	0,434	0,363	0,231	0,239	0,237	0,237
p_value	0,211	0,367	0,036	0,275	0,474	0,529	0,411	0,974	0,708	0,826	0,998	0,000
	A'	B'	C'	D'	E'	F'	G'	H'	I'	L'	M'	N'
RMSE	0,053	0,052	0,051	0,056	0,057	0,057	0,050	0,057	0,055	0,055	0,055	0,055
R^2	0,052	0,035	0,411	0,210	0,196	0,210	0,508	0,305	0,305	0,293	0,291	0,288
IA'	0,424	0,431	0,564	0,543	0,554	0,553	0,683	0,535	0,528	0,523	0,524	0,526
NSE	-0,086	-0,080	0,328	0,177	0,167	0,191	0,487	0,210	0,240	0,236	0,244	0,245
p_value	0,000	0,000	0,149	0,000	0,179	0,640	0,843	0,493	0,634	0,740	0,992	0,884

	PEAK_PH dataset	OFF-PEAK PH dataset	PEAK_wc dataset	OFF_PEAK_wc dataset
RMSE	0,7008	0,7466	0,0542	0,0552
R^2	0,9324	0,8728	0,3596	0,2218
IA'	0,8795	0,7311	0,6170	0,5341
NSE	0,9177	0,6523	0,3416	0,1817
p_value	0,7911	0,000013	0,5280	0,00000055

Table 17: Values of the set of metrics used to quantify the performance of the VGM in the different time periods ("peak periods" and "off-peak periods") using two different datasets (wc and PH dataset). The last group of metrics considers together all the peak periods (1st and 3rd columns) and all the off-peak periods (2nd and 4th columns) in order to give an overall indication of the performance of the model.

6_3 THE RELATION BETWEEN MODEL EVALUATION - MEASUREMENT ERRORS AND OTHER SOURCES OF UNCERTAINTIES

When dealing with model performance evaluation using observed data as in this case study, it is always important to keep in mind the existence of **measurement errors** that could partly undermine our evaluation. Measurements errors could assume different shape:

-Sensibility of the instruments to external and internal factors: in our case, GS3 sensors readings are sensible to temperature and bulk electrical conductivity while the dielectric water potential sensors (MPS6) and tensiometer T8 to temperature. These influences could not be avoided even with an optimal calibration.

-The error of the instruments recording the in situ measurements (water content and water potential). GS3 has an accuracy of +/- 0,03 m^3/m^3 while in our study, the variation between observed and simulated data is greater that this accuracy error (see Table 13, accuracy metrics for a wc dataset). For what concerns MPS6, the intrinsic accuracy of the instrument is pretty low, +/- 10% of the reading +2 kPa in the range from -9 to -100 kPa. This means for a reading of -100 kPa an estimated error between +12 kPa and -8 kPa which means +1,22 m and -0,81 m of water column height. T8 tensiometer has an accuracy significantly greater: +/-0,5 kPa (0,05 m of water column height) in the range -85 / +100 kPa. If we consider separately sensors T8-TB1-4,9 m (berm), MPS6-MPC1-4,6m (embankment), MPS6-SPC1-7m (embankment), the values of the statistical metrics are the following (Table 18) using the VGM:

	BEST FIT	T8_TB1 (m)	MPS6_MPC1_4,6m (m)	MPS6_SPC1_7m (m)
MBE	0	-0,87	0,35	-0,11
MAE	0	0,87	0,35	0,11
RMSE	0	1,22	0,40	0,23
R^2	1	0,98	1,00	1,00
NSE	1	0,96	1,00	1,00
IA	1	0,99	1,00	1,00
NSE_{rel}	1	-11,59	0,99	0,96
IA_{rel}	1	-2,76	1,00	0,99
NSE_j	1	0,82	0,96	0,98
IA_j	1	0,90	0,98	0,99

Table 18: Set of metrics used to investigate the performance of the VGM in the observation points T8_TB1; MCP1_4,6m and SPC1_7m.

Observation point T8-TB1 shows the greatest error both in accuracy and efficiency metrics but, according to what we have just exposed, T8-TB1 has also an higher nominal accuracy compared to MPS6. However it has to be considered that it was not installed during the INFRASAFE project but some time before by AIPO; for this reason the goodness of the installation was not verified and a lower reliability of the acquired data could not be excluded. This latter falls in ''other sources of uncertainties'', different from measurement errors. It is likely that even with an excellent calibration, the scatter between observed-simulated data remains high for the sensor T8_TB1 because observed data could be affected by intrinsic errors.

-Error correlated to the indirect method of measuring water content using the Capacitance Domain Reflectometry (CDR) as for GS3 sensor by Decagon Device.

-Instrument calibration is often obtained for just a portion of the SWRC, usually the wetter part, while the calibration of the dry part is usually problematic due to the difficulties in comparing instrument response with independent and accurate water potential/water content measurements. Good accuracy of instrument measurements in this narrow range is often theoretically achievable but not demonstrated for the whole range of soils. This is another important aspect to consider in errors estimation.

The following list falls into ''other sources of uncertainties'', different from measurement errors.

-Outliers in the observed dataset have a higher incidence on the evaluation of a model performance. For this reason, it is extremely important to minimize their number. If the number of observed data is too high to perform a manual screening, weighting techniques to optimize input data or probability based distributions (i.e. Monte Carlo simulation or Bayesian analysis framework) have to be used to deal with random or systematic errors (Phogat, 2016).

-Another possible source of scattering between observed and simulated dataset is the exact physical location of the instruments, since small changes in the depth could cause relatively large changes in water content and pore pressure values in time (Verbist et al, 2009). In the present case study, the sensors positioned in the boreholes have the accuracy of the centimetre and a displacement of a couple of centimeters is possible.

-An inaccurate representation of the initial conditions in terms of water content or suction distribution could be cause of the variability between modelled and observed data especially in the first time period of the simulation (as for GS3 SPC2 7,1 m and GS3 MPB2 2,2 m dataset in Figure 17).

-The natural variability intrinsic in the observation data to which the performance of the model is compared as soil heterogeneity that causes a dispersion of the hydraulic properties.

-Simplified assumptions made for lack of precise experimental data that cause an inaccurate description of physical phenomena.

For example, in Figure 21 various accuracy errors (MBE, MAE and RMSE), the coefficient of determination R^2 and the efficiency metrics (IA, KGE) are presented for the water content and pressure head datasets at different soil depths (from the most superficial observation point at the top, to the deepest at the bottom of the graph). It is possible to observe that the major bias is detected in the first 1,5 m of soil due to assumed modelling simplifications such as:

- The assumption of a constant atmospheric boundary flux during the whole day time neglecting the fluctuation in the evapo-transpiration contribution between day and night due to lack of experimental data.
- The weather station that provides meteorological data to compute the in-flow and out-flow from the atmospheric boundary is not positioned in the monitored section but in Cortile da Carpi (Lat. 44.778387, Lon. 10.971285), 7 km away from the investigated section, as reported in chapter §5.
- Other simplifications have been adopted for the estimation of the evaporation coefficient (K_e) and of the transpiration coefficient (K_{cb}) to compute ET_c, the actual crop evapotranspiration, because no proper information on the vegetative coverage of the investigated embankment are available.

For what concerns the pressure head dataset in the embankment below 1,5m, the coefficient of determination R^2 shows values always above 0,6 (best fit 1) and the accuracy errors (MBE, MAE, RMSE) decrease considerably up to 7m of depth. The accuracy metric IA follows the same trend of the accuracy metrics while KGE shows a decrement below 6,2m of depth. For what concerns the water content dataset, the observation points located in the embankment layer show an increase in the accuracy errors (MBE, MAE, RMSE) with depth (for the sensors below 1,5m), while R^2 does not have a clear trend but it is set to values always higher than 0,7. For what concerns the efficiency metrics IA and KGE, they show a decrease up to 4,5m of depth then a rapid increase (best fit value 1).

In the berm (fluvial layer), the values of the metrics suggest an higher discrepancy between observed and simulated datasets but with a coefficient of determination always greater than 0,5 (the value typically considered acceptable) with the exception of MPS6-MPB2-2,7m (R^2 of 0,38).

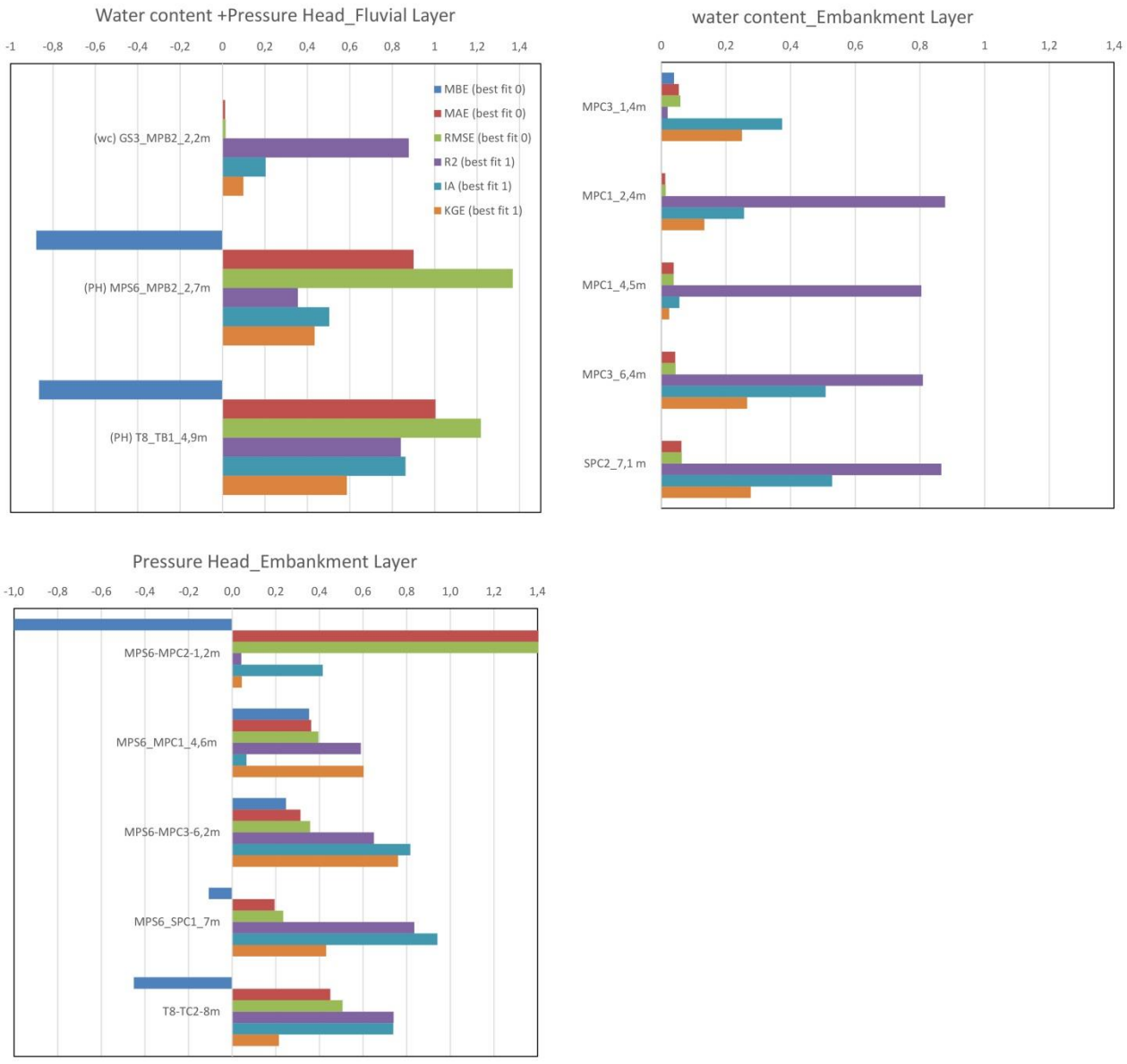


Figure 21: Values of the accuracy metrics (ME, MAE, RMSE, R^2) and the efficiency metrics (IA, KGE) with depth in the observation points located in the embankment layer and in the berm (fluvial layer). The shallower sensor is on the top of each graph and the deeper one at the bottom. The chosen hydraulic model is the VGM.

6_4 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The behaviour of the four considered hydraulic models (HVGM, VGM, VGM-2cm, HVGM-2cm) has been investigated in n°7 observation points belonging to different layers (embankment, fluvial, foundation, aquifer) using a PH, a wc dataset and a mixed dataset (PH+wc). The four performed direct simulations do not differ in any aspect (initial conditions, boundary conditions, mesh discretization etc) in order to be fully and easily compared. The comparison between the four different hydraulic models reveals if it is convenient to use a HVGM/HVGM-2cm model (10 parameters) instead of a VGM/VGM-2cm (6 parameters) for the investigated geotechnical problem. In fact the use of a more complex model requires the calibration of a higher number of parameters, longer calculation time and greater convergence problems, for this reason the choice has to be made on the basis of objective information.

Initially a qualitative procedure has been applied: the observed-simulated datasets have been represented on the same graph in order to visualize the dispersion from the 45° line (1:1) (see Figure 17 and Figure 18). Subsequently a set of metrics/indices has been used in order to quantify the performance of the model in a more objective way.

Using the qualitative methodology, it has been observed that VGM, HVGM, VGM-2cm show almost the same behaviour in the investigated observation points while VGM-2cm has a different trend. The considered models represent well the behaviour of the observation points (PH dataset) in the foundation and embankment layers and relatively well the behaviour in the embankment and fluvial layers (wc dataset). Comparing observed-simulated datasets for all the considered hydraulic models, it is possible to observe that the embankment is always less responsive to external stimuli with respect to models simulation. VGM and HVGM have the same response in the embankment layer but quite different in the berm. This is due to the fact that the duration of a flood event affects greatly the hysteretic behaviour of an embankment. If the permeability of the embankment is too small compared to the flood duration, the saturated front has not enough time to propagate deeply in the bank body and the resulting pore water pressure changes are small. The berm is the area which is primarily interested by the hydraulic stimuli and where the hydraulic stimuli are more persistent in time, for this reason the greater differences between the VGM and HVGM are observed in this area and in correspondence of the major floods events (for the hydrometric level reached and the persistence in time). It is relevant to consider these variations because the FOS (Factor of Safety) elaborated from proper stability analysis using as initial conditions the elaborated distributions of wc and pwp could result overestimated using a non-hysteretic behaviour as found by Liu et al (2015). The qualitative methodology to compare different hydraulic models, if in one hand results easier and quicker in order to have a general idea of the models performance, in the other it does not allow to elaborate a reliable and objective opinion, free from the personal judgement of the modeller. Metrics/indices could be used to evaluate the performance of the model/s on the single periods (on the single flood/off-flood periods), on the single observation point (local evaluation of the model performance) or globally on the whole domain (using the whole set of available observation points). The used metrics are the ones introduced in chapter 2_7_2 and tested in chapter 5 with similar datasets in order to individuate pros, cons and potential limits in their application. To this group of metrics, model selection criteria (AIC, BIC, HQC) have been added because extremely suitable for a probabilistic selection of the best model among many and because they are able to take into consideration the different complexity of each model. It has been observed that VGM and HVGM are the best models to simulate the 2018 period and HVGM is able to reproduce better the behaviour during the major flooding events. The following analysis concentrate the attention on the VGM which is the simpler hydraulic model used and whose performance has proven to be good. In this case to evaluate the model performance a greater number of observation points has been used as input data. As observed also in the previous analysis, the VGM gives an excellent performance in the simulation of the pwp distribution in time (values of the considered metrics close to the best fit) while an insufficient performance is observed for a wc dataset (below acceptability values). This tell us that the calibration is not going to enhance greatly the performance of the simulation in terms of pressure head, while greater room for improvements is possible for the simulation of the water content distribution. Subsequently the behaviour of the VGM has been investigated on the single period (flood/off-flood periods) by means of a restricted number of metrics/index in order to avoid a counter-productive number of data to be processed by the modeller. It is relevant to investigate the behaviour of the model expecially in the most relevant flood events of the 2018 period in view of possible future stability analyses. Even in this case the flood periods are well simulated compared to a PH dataset of observation points while

an insufficient performance is observed compared to a wc dataset. Chapter 6_3 focuses on the measurements errors that could partly influence the model evaluation. Among the most relevant ones, the sensitivity of the instruments to external/internal factors (as temperature, bulk electrical conductivity..), unfortunately those types of disturbance could not be avoided. Another important aspect to consider when dealing with data acquisition from installed instrumentations is the accuracy of thesensors for example GS3 has an high accuracy of +/- 0,03 m^3/m^3; T8 of +/-0,05m (in the range -85/+100 kPa) and MPS6 of -0,8/+1,22m (low degree of accuracy). Other sources of measurement errors stay in the laboratory technique to obtain the calibration curves of the instruments (f.e. the intrinsic difficulty in the experimental obtainment of the drying branch of the SWRC curve) and in the indirect method of measuring some properties/physical quantities (f.e. the wc is determined using the CDR technique for the sensor GS3). Other sources of uncertainties could be error in the vertical positioning of the sensors inside a borehole, soil local heterogeneity that causes a dispersion of the hydraulic parameters, the adoption of simplified assumption to describe physical phenomena (f.e evapo-transpiration phenomena) due to the lack of experimental proper data. When dealing with model performance evaluation using an observed dataset it is always suggested to understand the entity of the uncertainties with which the modeller is dealing in order to require to the parameters calibration process the reasonable degree of precision/accuracy.

The following chapter 7 focuses on the indirect simulations (inverse analysis) performed on the 2018 period in order to individuate the best set of optimized parameters. Indirect simulations have been carried out using different hydraulic models, different datasets of observation points, different weighting distributions, different set of initial parameters. A methodological approach in three phases has been applied in order to individuate the best set of optimized parameters. Indices/metrics/statistical parameters have been used for this purpose.

7 THE CALIBRATION PROGRAMME PERFORMED

In the present chapter, the performed calibration programme carried out on the direct simulations (HVGM, VGM, HVGM-2cm, VGM-2cm) of the 2018 year is presented in great detail.

The programme is composed of different **''groups of inverse analysis**'' to be performed, which are identified in Table 21 with the identification codes 1_1_xx with xx varying from 1 to 66.

As could be observed in Table 19, in each ''group of inverse analysis'' (denoted by the identification code 1_1_xx) 13 simulations are performed for the VGM and VGM-2cm while 18 simulations for the HVGM and HVGM-2cm. This is due to the fact that the VGM and VGM-2cm have a lower number of hydraulic parameters to optimize (lack of the hydraulic parameters that describe the wetting curve).

Each group of inverse analysis optimizes parameters of a different layer of the model (foundation, fluvial, embankment or aquifer) or of group of layers (embankment+fluvial; embankment+foundation; embankment+foundation+fluvial) (see Table 21).

The indirect simulations in each group of inverse analysis optimize one single parameter up to 5 parameters at a time when the goal is the optimization of a single layer (see Table 19). When a group of inverse analysis optimizes the hydraulic parameters of two or more layers, the indirect simulations tries to optimize from 2

to 15 hydraulic parameters (15 is the maximum number of parameters that could be optimized together in the Hydrus code).

	HVGM – HVGM-2cm									VGM - VGM-2cm				
	θr	θs	α	n	K_s	θ_m	θ_{sw}	α_w	k_{sw}	θ_r	θ_s	α	n	K_s
1					X									X
2				X									X	
3		X	X									X		
4	X										X			
5				X	X					X				
6			X		X							X		X
7		X			X						X			X
8	X				X					X				X
9		X	X		X						X	X		X
10	X			X	X					X			X	X
11		X		X	X						X		X	X
12		X	X	X							X	X	X	
13	X	X	X	X	X					X	X	X	X	X
14						X								
15							X							
16								X						
17									X					
18						X	X	X	X					

Table 19: The table represents the optimized parameters in each indirect simulation belonging to a group of inverse analysis (denoted by the identification code 1_1_xx). Each group of inverse analysis is composed of 18 indirect simulations (numbered from 1 to 18 in the first column) for the HVGM and HVGM-2cm, of 13 indirect simulations for the VGM and VGM-2cm (because the hydraulic parameters of the wetting curve are not to be optimized).

Each group of inverse analysis has different input dataset to perform the inverse modelling: pressure head data or/and water content data of the sensors installed in the layer/s whose parameters are optimized (see Table 21). In Table 20, the sensors used for the inverse modelling are listed. When the inverse analysis is performed on a group of layers, the observation dataset for the indirect problem is composed of pressure head or/and water content data of the available instruments belonging to the group of layers considered.

In other words, if the group of inverse analysis optimizes the hydraulic parameters of the embankment layer using a water content dataset, only the observation points/sensors monitoring the water content in the investigated layer are used in the inverse analysis. If the group of inverse analysis optimized the hydraulic parameters of two layers (f.e. embankment and fluvial) using a pressure head (PH) dataset only the sensors that records pwp installed in the two investigated layers are used in the indirect problem. This concept is clearly visible in column 8 of Table 21 where the groups of inverse analysis performed in the present calibration programme are listed reporting information about type of dataset used, sensors used in the dataset of observation points, weighting distributions, hydraulic model adopted, layer/s whose parameters are optimized.

The comparison of the performance of indirect simulations belonging to different groups of inverse analysis (as the one performed in the first phase of the calibration) is carried out considering the whole set of observation points (all the available sensors) and not only the restricted number of sensors used in the inverse analysis. This assures a comparison of the indirect simulations under the same conditions.

DATASET IN PRESSURE HEAD (m) - EMBANKMENT	DATASET IN WATER CONTENT (%) – EMBANKMENT
T8-TC2-8,0m	GS3-SPC2-7,1m
MPS6-MPC1-4,6m	GS3-MPC1-4,5m
MPS6-SPC1-7m	GS3-MPC1-2,4m
MPS6-MPC3-6,2m	GS3-MPC3-6,4m
MPC1-3,05m	
DATASET IN PRESSURE HEAD (m) - FOUNDATION	**DATASET IN WATER CONTENT (%) – FLUVIAL**
T8-TB1-4,9m	GS3-MPB2-2,2m
PZ1-10m	**DATASET IN PRESSURE HEAD (m) – AQUIFER**
MPS6-MPB2-2,7m	PZ2-17m

Table 20: Sensors used to perform the inverse modelling subdivided for type of measure (wc or PH) and layer.

A simplified example is provided below in order to clarify how the calibration programme has been conceived and carried out. Assuming to have:

- two layers (A,B) whose hydraulic behaviour is described by 2 parameters each: A (a,b) and B (c,d),

-a unique weighting distribution (w=1);

-two types of datasets of observation points (X and Y);

All the groups of inverse analysis that are performed according to the calibration programme adopted are presented in Figure 22. 6 groups of inverse analysis: 4 groups optimizing a single layer (A or B) and 2 optimizing both the layers (A+B), what changes is the dataset of the observation points (X or Y) because the weighting distribution is unique in this example. In Figure 22 for each group of inverse analysis, possible indirect simulations are reported.

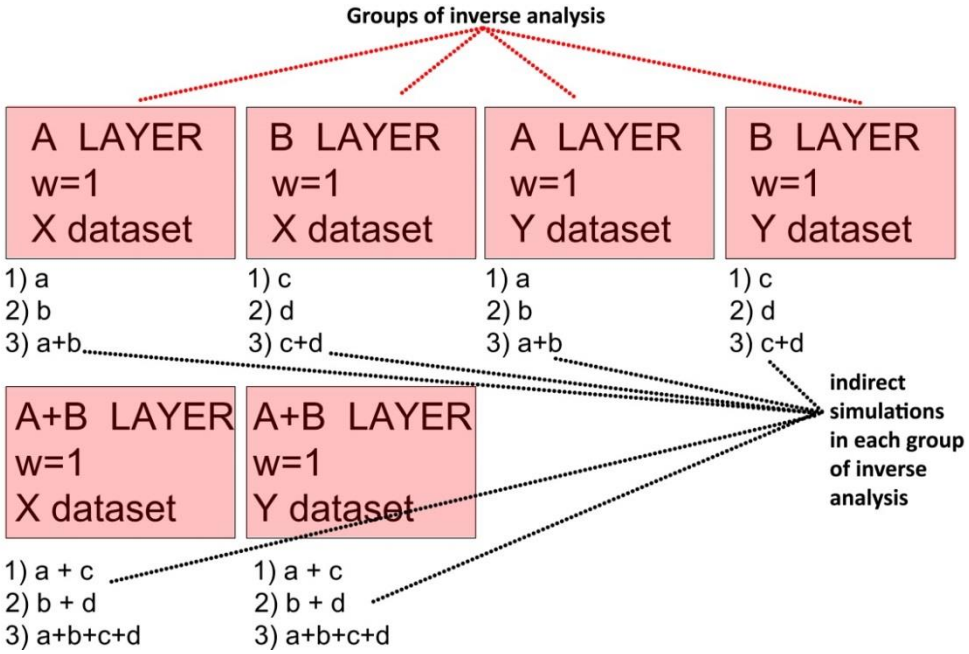


Figure 22: Example of the groups of inverse analysis that ARE performed according to the calibration programme adopted. In this example we assume to have 2 layers whose parameters have to be optimized, a single weighting distribution (w=1) and two types of dataset of observation points (X and Y).

The instruments installed in the first 1,5m from the soil surface have not been included in the inverse analysis due to the high influence of the evapo-transpiration contribution that has been evaluated with a low level of accuracy (no in situ weather station, no information to estimate precisely the crop coefficient of the riparian vegetation and the evaporation coefficient Ke, no hourly computation of the evapotranspiration contribution but only daily computation). This topic has been addressed with due care in the previous chapter §6_3. The sensors that have not been included in the indirect modelling are the following: GS3-MPB1-0,7m (wc), GS3-MPC3-1,4m (wc), MPS6-MPB1-0,9m (PH), SM150T-MPC2-1,5m (PH) and MPS6-SPB1-1,2m (PH).

The observation data used in the inverse modelling have been associated to different weights:

-a weight equal to 1 for the whole observation dataset (the same relevance is assumed for the whole set of observation points).

-a weight associated to input data according to the Triangular Moving Average Technique (TMAT).

-a more complete casuistry on the attribution of weights to the input data on the basis of the type of the instrument (reliability and quality of the readings), of the position of the instrument in the bank, of the temporal instant of the observation data (pre, ongoing, after flood, no flood) will be presented separately and extensively in chapter §7_7. The different weighting procedures have been applied to a reduced group of indirect simulations.

The Triangular Moving Average Technique (TMAT) was born for the financial market. It is a simple moving average that is averaged again (averaged twice) so it is double smoother. It is not a responsive type of moving average (the reaction to a sudden change in the dataset is not very fast). TMAT is formally defined the weighted average of the last ''n'' values and of the future ''m'' values of a dataset (x_i with i=1..n+m), and it could be computed as follows [see equations 129 and 130]:

SMA=single moving average= $(x_1+ x_2+….+x_{n+m})/(n+m)$ [equation 129]

TMA= $(SMA_1+SMA_2+…..SMA_{n+m})/(n+m)$ [equation 130]

In the TMAT, the majority of the weights is given to the middle portion of the data of the look-back/look-ahead period and the weight decreases moving away from the middle as in a triangle, hence the name. In our case, the look-back period is one week and the look-ahead period is one week, so in total the period to which the TMA is referred to is two weeks. The max weight is given to the peak value of the instrument (i.e. during a flood event), and the rest of the data is scaled from this value using the TMAT.

The TMAT has been chosen because it allows to give sufficient weight also to the data belonging to the initial phase and post peak phase of a flood event because they are relevant for the closeness to the peak. These two temporal instants are particularly interesting especially from the point of view of future stability analysis. In fact a sudden high water level is potentially dangerous for medium-coarse grained banks and for this reason great accuracy must be dedicated to represent the pore pressure distribution in the first days of the flood, while a persistent medium/high water level is potentially dangerous for fine grained banks and for this reason the ''off-peak'' of the flood (last days of the flood) must be well defined in terms of pressure head distribution. As could be observed in Table 21, the parameters to be optimized have been constrained (setting the range in which the parameter could vary) just in the simulations groups 1_1_1 and 1_1_2 due to the fact that the optimized parameters have always assumed possible values. In none of the

other performed groups of inverse analysis a constrain was necessary. For this reason, the present case study could not be used to investigate the effects of constrained and unconstrained optimized parameters.

Concluding, the different characteristics of the indirect simulations carried out in the performed calibration programme are summed in Figure 23 for sake of clarity.

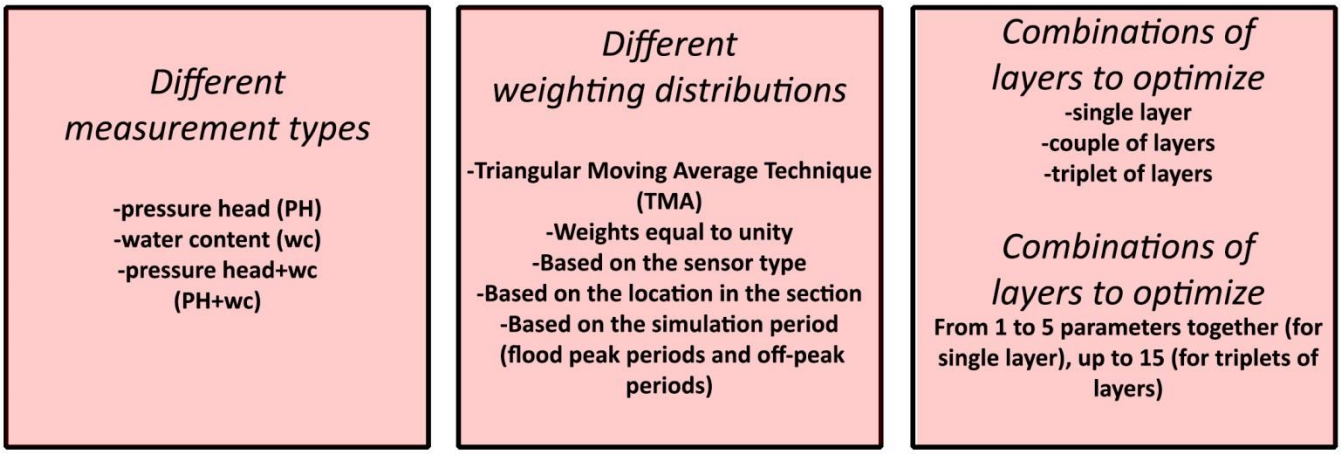


Figure 23: Summary of the different characteristics of the indirect simulations carried out in the performed calibration programme.

INVERSE ANALYSIS GROUP	HYDRAULIC MODEL	CONSTRAINED PARAMETER/S	UNCOSTRAINED PARAMETER/S	INPUT DATA: PRESSURE HEAD	INPUT DATA: wc	LAYER/S TO WHICH OPTIMIZED PARAMETERS BELONG	LAYER/S TO WHICH INPUT DATA BELONG	WEIGHTS OF THE OBSERVED DATA=1	WEIGHTS OF THE OBSERVED DATA=TMA
1_1_1	HVGM	X		X		EMBANKMENT	EMBANKMENT	X	
1_1_2	HVGM	X			X	EMBANKMENT	EMBANKMENT	X	
1_1_3	HVGM		X	X		EMBANKMENT	EMBANKMENT	X	
1_1_4	HVGM		X		X	EMBANKMENT	EMBANKMENT	X	
1_1_5	VGM		X	X		EMBANKMENT	EMBANKMENT	X	
1_1_6	VGM		X		X	EMBANKMENT	EMBANKMENT	X	
1_1_7	HVGM		X		X	FLUVIAL	FLUVIAL	X	
1_1_8	VGM		X		X	FLUVIAL	FLUVIAL	X	
1_1_9	HVGM		X	X		FOUNDATION	FOUNDATION	X	
1_1_10	VGM		X	X		FOUNDATION	FOUNDATION	X	
1_1_11	HVGM		X		X	EMB+FLUV	EMB+FLUV	X	
1_1_12	VGM		X		X	EMB+FLUV	EMB+FLUV	X	
1_1_13	HVGM		X	X		EMB+FOUND	EMB+FOUND	X	
1_1_14	VGM		X	X		EMB+FOUND	EMB+FOUND	X	
1_1_15	HVGM		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV	X	
1_1_16	VGM		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV	X	
1_1_17	HVGM		X	X		AQUIFER	AQUIFER	X	
1_1_18	VGM		X	X		AQUIFER	AQUIFER	X	
1_1_19	HVGM-2cm		X	X		EMBANKMENT	EMBANKMENT	X	
1_1_20	HVGM-2cm		X		X	EMBANKMENT	EMBANKMENT	X	
1_1_21	VGM-2cm		X	X		EMBANKMENT	EMBANKMENT	X	
1_1_22	VGM-2cm		X		X	EMBANKMENT	EMBANKMENT	X	
1_1_23	HVGM-2cm		X		X	FLUVIAL	FLUVIAL	X	
1_1_24	VGM-2cm		X		X	FLUVIAL	FLUVIAL	X	
1_1_25	HVGM-2cm		X	X		FOUNDATION	FOUNDATION	X	
1_1_26	VGM-2cm		X	X		FOUNDATION	FOUNDATION	X	
1_1_27	HVGM-2cm		X		X	EMB+FLUV	EMB+FLUV	X	
1_1_28	VGM-2cm		X		X	EMB+FLUV	EMB+FLUV	X	
1_1_29	HVGM-2cm		X	X		EMB+FOUND	EMB+FOUND	X	
1_1_30	VGM-2cm		X	X		EMB+FOUND	EMB+FOUND	X	
1_1_31	HVGM-2cm		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV	X	
1_1_32	VGM-2cm		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV	X	
1_1_33	HVGM-2cm		X	X		AQUIFER	AQUIFER	X	
1_1_34	VGM-2cm		X	X		AQUIFER	AQUIFER	X	
1_1_35	HVGM		X	X		EMBANKMENT	EMBANKMENT		X
1_1_36	HVGM		X		X	EMBANKMENT	EMBANKMENT		X
1_1_37	VGM		X	X		EMBANKMENT	EMBANKMENT		X
1_1_38	VGM		X		X	EMBANKMENT	EMBANKMENT		X
1_1_39	HVGM		X		X	FLUVIAL	FLUVIAL		X
1_1_40	VGM		X		X	FLUVIAL	FLUVIAL		X
1_1_41	HVGM		X	X		FOUNDATION	FOUNDATION		X
1_1_42	VGM		X	X		FOUNDATION	FOUNDATION		X
1_1_43	HVGM		X		X	EMB+FLUV	EMB+FLUV		X
1_1_44	VGM		X		X	EMB+FLUV	EMB+FLUV		X
1_1_45	HVGM		X	X		EMB+FOUND	EMB+FOUND		X
1_1_46	VGM		X	X		EMB+FOUND	EMB+FOUND		X
1_1_47	HVGM		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV		X
1_1_48	VGM		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV		X
1_1_49	HVGM		X	X		AQUIFER	AQUIFER		X
1_1_50	VGM		X	X		AQUIFER	AQUIFER		X
1_1_51	HVGM-2cm		X	X		EMBANKMENT	EMBANKMENT		X
1_1_52	HVGM-2cm		X		X	EMBANKMENT	EMBANKMENT		X
1_1_53	VGM-2cm		X	X		EMBANKMENT	EMBANKMENT		X
1_1_54	VGM-2cm		X		X	EMBANKMENT	EMBANKMENT		X
1_1_55	HVGM-2cm		X		X	FLUVIAL	FLUVIAL		X
1_1_56	VGM-2cm		X		X	FLUVIAL	FLUVIAL		X
1_1_57	HVGM-2cm		X	X		FOUNDATION	FOUNDATION		X
1_1_58	VGM-2cm		X	X		FOUNDATION	FOUNDATION		X
1_1_59	HVGM-2cm		X		X	EMB+FLUV	EMB+FLUV		X
1_1_60	VGM-2cm		X		X	EMB+FLUV	EMB+FLUV		X
1_1_61	HVGM-2cm		X	X		EMB+FOUND	EMB+FOUND		X
1_1_62	VGM-2cm		X	X		EMB+FOUND	EMB+FOUND		X
1_1_63	HVGM-2cm		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV		X
1_1_64	VGM-2cm		X	X	X	EMB+FOUND+FLUV	EMB+FOUND+FLUV		X
1_1_65	HVGM-2cm		X	X		AQUIFER	AQUIFER		X
1_1_66	VGM-2cm		X	X		AQUIFER	AQUIFER		X

Table 21: Summary of the characteristics of each group of inverse simulations (numbered from 1_1_1 to 1_1_66 in the first column) in terms of hydraulic model adopted (VGM, HVGM, VGM-2cm, HVGM-2cm), type of observation dataset (wc and/or

PH); layer/s to which optimized parameters belong; layer/s to which observation dataset belongs, weighting technique adopted for the observation dataset.

7_1 PRELIMINARY ANALYSIS OF THE RESULTS OF THE CALIBRATION PROGRAMME

				% OF SUCCESSFUL SIMULATIONS OPTIMIZING:				
	HYDRAULIC MODEL	n° OF INDIRECT SIMULATIONS PERFORMED FOR EACH LAYER	% OF SUCCESSFUL SIMULATIONS([1])	1 PARAM.([2])	2 PARAM.([3])	3 PARAM.([4])	4 PARAM.([5])	5 PARAM.([6])
EMBANKMENT LAYER	HVGM	108	55	56	63	38	67	50
	VGM	52	46	55	55	25	50	50
	VGM-2cm	52	12	27	0	0	25	25
	HVGM-2cm	72	0	0	0	0	0	0
FLUVIAL LAYER	HVGM	36	29	38	30	0	0	0
	VGM	26	27	30	30	50	0	50
	VGM-2cm	26	19	30	10	0	50	0
	HVGM-2cm	36	0	0	0	0	0	0
FOUNDATION LAYER	HVGM	36	47	38	80	25	0	0
	VGM	26	69	60	80	100	50	50
	VGM-2cm	26	15	40	0	0	0	0
	HVGM-2cm	36	18	25	20	0	0	0
AQUIFER LAYER	HVGM	36	53	31	80	50	100	100
	VGM	26	92	80	100	100	100	100
	VGM-2cm	26	58	10	80	100	100	100
	HVGM-2cm	36	0	0	0	0	0	0
EMBANKMENT +FLUVIAL LAYERS	HVGM	36	18	25	10	25	0	0
	VGM	26	15	30	0	0	50	0
	VGM-2cm	26	4	0	10	0	0	0
	HVGM-2cm	36	0	0	0	0	0	0
EMBANKMENT +FOUNDATION LAYERS	HVGM	36	35	38	50	25	0	0
	VGM	26	35	60	30	0	0	0
	VGM-2cm	26	4	0	10	0	0	0
	HVGM-2cm	36	0	0	0	0	0	0
EMBANKMENT+ FOUNDATION+ FLUVIAL LAYERS	HVGM	36	21	30	10	0	0	0
	VGM	26	35	50	30	50	0	0
	VGM-2cm	26	0	0	0	0	0	0
	HVGM-2cm	36	0	0	0	0	0	0
Total n° of indirect simulations		1028	256					

Table 22: The table reports in the third column the number of performed simulations subdivided according to different layer (or group of layers) and the different hydraulic models; in the fourth column the number of successful simulation for each group, from the fifth to the ninth column the % of successful simulations subdivided according to the number of optimized parameters (1 to 5 parameters)

([1]) with respect to the total number of simulations performed for each layer and for each hydraulic model (reported in the second column). (2) with respect to the total number of performed simulations that optimize one parameter, (3) two parameters, (4) three parameters, (5) four parameters, (6) five parameters.

From Table 22 the 66 groups of inverse analysis are divided as follows:

- 18 groups optimize the hydraulic parameters of the embankment layer
- 8 groups optimize the hydraulic parameters of the fluvial layer
- 8 groups optimize the hydraulic parameters of the foundation layer
- 8 groups optimize the hydraulic parameters of the embankment + fluvial layers
- 8 groups optimize the hydraulic parameters of the embankment + foundation layers
- 8 groups optimize the hydraulic parameters of the embankment + fluvial + foundation layers
- 8 groups optimize the hydraulic parameters of aquifer layer

The starting 66 groups of inverse analysis (first column of Table 21) reduce to 33 excluding the groups with any successful simulations (no convergent simulations).

In Table 22, the number of indirect simulations performed for each layer and for each hydraulic model (VGM, HVGM, VGM-2cm, HVGM-2cm) is presented in the first column. The reported number of simulations for each hydraulic model refers to the simulations that use a pressure head dataset, a water content dataset (or both) and the two weighting distributions (weights equal to unity and weights distributed according to the TMAT).
The total number of indirect simulations performed in the calibration phase is 1028 (324 with HVGM, 208 with VGM, 208 with VGM-2cm, 288 with HVGM-2cm), while the number of successful indirect simulation is 256. Only the 25% of the initial set of indirect simulations achieved convergence. With high probability this was a consequence of the complexity of the problem considering a multi-layered domain, and a medium-long period of simulation.
For the embankment layer, a greater number of simulations was performed with respect to the other layers because both water content and pressure head observation data were available within the same layer. Moreover the indirect simulations of the groups of inverse analysis 1_1_3 and 1_1_4 were redone with constrains in the optimized parameters (1_1_1 and 1_1_2).
Constraining a variable means to specify an upper and lower bound so the objective function, by varying the optimization variable within these bounds, will dismiss any solutions which do not meet the specified constraint condition.
Looking at the results of the optimization, there was no need to impose constrains to the hydraulic parameters of any layer of the river bank section.
In the fourth column of Table 22, the percentages of convergent indirect simulations (with respect to the total number of simulations indicated in the second column) are reported. It is relevant to stress that the same calibration procedure has been applied to each layer[2], including the aquifer for which the optimization of all the hydraulic parameters was performed despite the insensitivity of almost all the parameters (the only exception is the saturated permeability). Therefore the high values of successful simulations of the aquifer layer are not relevant.
As could be observed in the third column of Table 22, the VGM and HVGM secure an high percentage of successful inverse simulations for the embankment layer (>50%) and for the foundation layers (69%), while, for the fluvial layer, HVGM and VGM assure the same low percentage of successful simulations (27-29%).

HVGM-2cm shows a very low percentage of success for all the layers (0 or less than 1%) so we can conclude that it is impossible to calibrate the parameters of this model by inverse analysis. One of the most probable reason may be found in the ill-posed inverse problem. In a complex n-layer and multi-parametric model many factors contribute to inverse-analysis nonuniqueness as indeterminacy, statistical data-error distribution, numerical error and instability, finite data and model parameter (Simunek and Hopmans, 2002).
It is well known that the use of additional data in the indirect problem together with the reduction of the number of parameters involved in the optimization process could reduce or resolve these problems. Both strategies have been tested in the calibration programme but no improvements have been observed. For what concerns the VGM-2cm the percentage of successful simulations is low (from 0 to 19%).

From the fifth column to the ninth column (Table 22), the percentage of successful indirect simulations optimizing from 1 to 5 parameters is reported based on the initial number of simulation. The number of optimized parameters is an important factor affecting the inverse modelling. In fact local gradient based optimization algorithm such as the Levenberg-Marquardt method implemented in Hydrus works best with a limited number of parameters. If a large number of parameters needs to be calibrated, other optimization algorithm such as the Genetic Algorithms are to be preferred (Simunek et al, 2012; Vrugt et al, 2001). The

[2] The same calibration programme has been applied to all the investigated layers (embankment, fluvial, foundation, aquifer) in order to assure an easier comparison between the obtained results. A reduction of the number of indirect simulations could be obtained optimizing only the hydraulic parameters which showed the greater sensitivity in chapter 6_1 (sensitivity analysis) but for the purposes of the present thesis, it was not a priority to maintain a low number of inverse analysis.

percentage of successful simulations seems not to be directly related to the number of optimized parameters (f.e. the embankment layer has a percentage of convergence equal to 56% for 1 parameter, 63% for two parameters and 67% for four parameters) but we are not considering any information on the quality of the inverse simulations moving from 1 to 5 optimized parameters (simulations that optimize more parameters are most likely less accurate than those optimizing fewer parameters).

While in Table 22 the subdivision between indirect simulations has been performed according to the different optimized layers and the different hydraulic models; in Table 23:

- All the successful simulations that use the TMAT are classified according to the different hydraulic models (case A).
- All the successful simulations that use weights equal to 1 are classified according to the different hydraulic models (case B).
- All the successful simulations are classified according to the different hydraulic models and the different type of dataset (wc and/or PH) (case C).
- All the successful simulations (optimizing the hydraulic parameters of each layer) are classified solely according to the different hydraulic models (case D).

As it is possible to observe, case A has a lower percentage of successful simulations with respect to case B. This is due to the fact that, in case A, an higher weight is given to observations belonging to flood events (pre, ongoing, post) and this implies an higher difficulty to minimize the objective function and find an optimized set of parameters. The algorithm tries harder to reduce the residuals between simulated and observed datasets where the weight of the dataset is higher (during the flood events) while the remaining periods affect less significantly the inverse simulation. Undoubtedly, it is easier to find a set of parameters that enhances the performance of the model during the ''off-peak'' periods (low river hydraulic level), more complex is the improvement of the simulation behaviour during the peak periods.
In table 20, as in table 19, there is no evidence of any correlation between the percentage of successful simulations and the increase in the number of optimized parameters.
VGM is the hydraulic model that shows the highest percentage of successful indirect simulations among the considered ones (46%, case D) and this number is kept high also for simulations with more than one optimized parameter while, as seen in Table 22, VGM-2cm has a low number of converged simulations (14%) and HVGM-2cm a percentage close to zero.
HVGM, VGM and HVGM-2cm with a pressure head dataset (case C) have the higher percentage of successful simulations compared to simulations with a water content dataset or a mixed dataset (PH+wc). For the purposes of this calibration procedure, a good performance of the model in terms of pore pressure distribution in time is preferable to the one in terms of w.c. distribution. The use of a w.c. dataset assures the lowest percentage of converged simulations while the mixed dataset (wc+PH) a value in between.

			% OF SUCCESSFUL SIMULATIONS OPTIMIZING				
	HYDRAULIC MODEL	% OF SUCCESSFUL SIMULATIONS ([1])	1 PARAM. ([2])	2 PARAM. ([3])	3 PARAM. ([4])	4 PARAM. ([5])	5 PARAM. ([6])
TMAT [case A]	HVGM	32	31	33	35	14	25
	VGM	32	38	30	25	38	25
	VGM-2cm	2	0	0	0	0	0
	HVGM-2cm	15	15	13	25	50	25
Weights equal to 1 [case B]	HVGM	49	50	63	25	38	30
	VGM	60	68	60	63	38	50
	VGM-2cm	2	3	3	0	0	0
	HVGM-2cm	6	14	15	8	13	13
Subdivision based on the hydraulic	HVGM (wc)	32	44	27	18	0	13
	HVGM (PH)	44	36	60	39	33	33
	HVGM (PH+wc)	21	38	10	0	0	0

model and the type of observation dataset [case C]	VGM (wc)	28	40	23	17	33	33
	VGM (PH)	57	52	65	50	50	50
	VGM (wc+PH)	35	50	30	50	0	0
	VGM-2cm (wc)	15	28	7	0	33	17
	VGM-2cm (PH)	17	8	23	25	25	25
	VGM-2cm (PH+wc)	0	0	0	0	0	0
All layers together [case D]	HVGM	41	42	49	29	27	28
	VGM	46	53	45	44	38	38
	HVGM-2cm	1	2	1	0	0	0
	VGM-2cm	14	12	14	11	25	19

Table 23: In the third column, the percentage of successful simulations in different cases (from A to D) subdivided according to the hydraulic models considered. From the fourth to eighth column, the percentage of successful simulations subdivided according to the number of optimized parameters.

([1]) with respect to the total number of simulations performed for each hydraulic model. (2) with respect to the total number of performed simulations that optimize one parameter, (3) two parameters, (4) three parameters, (5) four parameters, (6) five parameters.

7_1_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

Chapter 7_1 presents the results of the inverse analysis performed on the 2018 period. N°1028 indirect simulations have been carried out and they could be grouped in n° 66 ''groups of inverse analysis''. N° 324 indirect simulations have been performed using the HVGM, n°208 the VGM, n°208 the VGM-2cm and n°288 the HVGM-2cm. Each group of inverse analysis contains 18 simulations for HVGM and HVGM-2cm and 13 simulations for VGM and VGM-2cm. Each group of inverse analysis uses a different dataset of observation points, a different hydraulic model (VGM, VGM-2cm, HVGM, HVGM-2cm), optimizes different layers (single, couple, triplets of layers) and different combinations of hydraulic parameters (from 1 to 5 for the optimization of a single layer up to 15 parameters for triplets of layers). The weighting distributions adopted are: all observed data equal to unity, weights distributed according to the Triangular Moving Average Technique (TMAT), weights attributed depending on the typology of sensors, the position of the sensor in the section, the time period in which data have been recorded, the reliability and quality of the data recorded by the sensor. It has been observed that the performed inverse analyses do not need constraints on the hydraulic parameters which vary in all the cases in a plausible range. This means that the investigated problem is not appropriate to deal with this topic (influence of constrains on the inverse analysis results). Only the 25% of the initial simulations have reached convergence (256) and, from 66 groups of inverse analysis, only the performance of 33 groups will be tested in the following chapters applying a methodological approach in three phases. With high probability, the low percentage of successful simulations is due to the complexity of the problem we are dealing with (long period of simulation, multi-layered domain, multi-parametrical hydraulic models). The VGM and HVGM secure an high percentage of successful inverse simulations for the embankment layer (>50%), for the foundation (69%) and aquifer (> 50%) layers, while, for the fluvial layer, HVGM and VGM assure the same low percentage of successful simulations (27-29%). HVGM-2cm shows a very low percentage of success for all the investigated layers (close to 0) due, with high probability, to an ill-posed inverse problem. Attempts have been made to eliminate or reduce this problem, decreasing for example the number of optimized parameters and incrementing the observation data used for the inverse analysis but with no success. We can conclude that the optimization of the hydraulic parameters of HVGM-2cm is not possible in the problem under analysis. The VGM-2cm shows, at the same way, a low number of successful indirect simulations with a percentage between 0-19% for all the investigated layers. Despite this, an attempt has

been made to use the available successful simulations to optimize the hydraulic parameters of this model. The VGM shows the highest percentage of successful simulation with respect to HVGM, HVGM-2cm, VGM-2cm if compared to the initial number of performed simulations. This is easily explained by the fact that the VGM is the easier model used which requires the lower number of parameters. The correlation between number of optimized parameters and the percentage of successful simulations has been investigated. A direct correlation is not visible passing from 1 to 5 optimized parameters in the four different hydraulic models. From literature, it is well known that the Levenberg-Marquardt algorithm works best with a limited number of parameters, this helps to reduce convergence problems, problems of non-uniqueness of the solution and ill-posedness of the inverse problem. In the present case study, a reduction of the percentage of the successful simulations together with a reduction of the number of optimized parameters is not observed for all the different hydraulic models. This means that it is more relevant to consider the quality of the simulations increasing the number of parameters instead of considering the percentage of successful simulations as it is expected a reduction of the improvement in the simulation output incrementing the number of optimized parameters. It has been observed an higher percentage of successful simulations using a weighting distribution with all weights equal to 1 with respect to a TMA weighting distribution. This is due to the fact that the TMA weighting distribution gives more relevance to the observation data during the peak periods and decrementally moving away from the peaks. In this case, the minimization of the objective function is more difficult because, if we concentrate the attention on the flood-events, it is particularly complex the individuation of a set of optimized parameters that enhance the model performance. On the contrary using a weighting distribution w=1 for the whole dataset of observation points, the inverse analysis is influenced mainly by the off-peak periods and the individuation of a set of optimized parameters results easier. Despite this, the optimization of the peak periods results more relevant for the purposes of our analysis (stability analyses in the most relevant instants of the simulation). It has been observed that for all the considered hydraulic models, the pressure head dataset assures the higher percentage of successful simulations with respect to a wc dataset or a mixed dataset (PH+wc). The wc dataset is associated to the lowest number of successful simulations while the mixed dataset (wc+PH) has a value in between.

While in chapter 7_1 the investigation involves the relation between the number of successful simulations and the characteristics of the indirect problem (number of optimized parameters, adopted hydraulic model, typology of dataset etc), in chapter 7_2 the focus is on the quality of the performance of the groups of inverse analysis in fact the match between observed-simulated datasets is investigated by means of metrics/indices.

7_2 PRELIMINARY EVALUATION OF THE PERFORMANCE OF GROUPS OF INVERSE ANALYSIS

In the present chapter §7_2, a preliminary evaluation of the performance of the **groups of inverse analysis** (with the identification code 1_1_xx with xx varying from 1 to 66) is presented, while in chapters §7_3, §7_4 and §7_5, the quantitative evaluation of the performance is concentrate on the **single indirect simulations** belonging to the different groups of inverse analysis.

The performance of a group of inverse analysis is evaluated by means of a set of metrics. The metrics are computed for each simulation belonging to the group of inverse analysis under examination, then an average value will be associated to the whole group.

The average could be considered significant to represent the behaviour of a whole group of inverse analysis because every simulation (which belongs to a certain group) optimizes the hydraulic parameters of the same layer, with the same dataset of observation points and the same distribution of the weights. We have noted that a metric computed on the output of all simulations belonging to a group of inverse analysis, is within a very narrow range. This makes legitimate and significant to consider the mean of these values as representative of the behaviour of a certain group of inverse analysis.

All the available groups of inverse analysis have been considered (33 groups). The proposed preliminary evaluation is suggested when a large number of groups is available, at the very initial phase of the calibration process, when it is necessary to look globally at the behaviour of the model excluding groups of inverse analysis that do not comply a minimum threshold that we are going to define. This preliminary phase is important in order to reduce the number of groups that will be investigated in detail in the next phases, excluding the ones that perform poorly. The use of threshold values for the considered set of metrics and the use of a colour scale to detect which groups of inverse analysis fulfil or not the required metrics allow an easy and rapid evaluation. Having an idea of the general trend of the simulations, allows to detect the correct thresholds and the scale of scores to adopt in the next phase where we look at the single simulations one by one investigating the performance not only during the whole simulation period (2018 year) but also during flood peak and off-peak periods.

The performance during the ''whole simulation period'' (2018 year) of each group of inverse analysis has been investigated by means of the following set of metrics:

-**Efficiency criteria**: the Nash-Sutcliffe Efficiency Criteria (NSE), the Index of Agreement (IA), the Modified Index of Agreement (IA'), the relative Nash-Sutcliffe Efficiency Criteria (NSE_{rel}), the Relative Index of Agreement (IA_{rel}), the modified Index of Agreement (IA_j) with j=1; the modified Nash-Sutcliffe Efficiency Criteria (NSE_j) with j=1, the Kling-Gupta Efficiency Criteria (KGE), the Coefficient of Determination (R^2).
-**Accuracy criteria**: the Mean Bias Error (MBE), the Mean Absolute Error (MAE), the Coefficient of Residual Mass (CRM), the Mean Square Error (MSE).
-**Model Selection Criteria**: the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), the Hannan-Quin Information Criterion (HQIC).

The mathematical and critical discussion of these parameters could be found in chapter §2_7_2. The threshold values adopted for each metric/index are the following:

- R^2 :values between 0,5 and 1 (1 is the best fit).
- NSE: values between -1 and 1 (1 is the best fit).
- NSE_{rel}: values between -1 and 1 (1 is the best fit).
- NSE_j: values between -1 and 1 (1 is the best fit).
- IA: values between 0 and 1 (1 is the best fit).
- IA_{rel}: values between 0 and 1 (1 is the best fit).
- IA_j: values between 0 and 1 (1 is the best fit).
- IA': values between 0 and 1 (1 is the best fit).
- KGE : values between -1 and 1 (1 best fit).
- MBE: values between -0,5 and +0,5 (0 best fit).
- MAE: values between 0 and +0,7 (0 best fit).
- CRM: values between -1 and +1 (0 best fit).
- MSE: values between -0,7 and +0,7 (0 best fit).
- AIC/AIC_{min}: values between 0 and 10 (0 best fit).
- BIC/BIC_{min}: values between 0 and 10 (0 best fit).
- $HQIC/HQIC_{min}$: values between 0 and 10 (0 best fit).

In chapter §4 the performance of all the chosen indicators has been investigated. It has been observed that IA, IA_{rel} and IA_j in some cases show an inconsistent behaviour that lead to prefer the modified form IA'. Inconsistencies were observed also in some cases for KGE, CRM and R^2. The pros and cons of these indices have been presented extensively in chapter §2_7_2 and, as said many times, their combined use is the key for a good simulation performance evaluation. In this preliminary phase, all these indices (IA, IA_{rel}, IA_j, KGE, CRM and R^2) have been considered because they are able to catch the global behaviour of the groups of inverse analysis, while their use for a hierarchical classification of the groups/simulations (from the better to the worst) is not recommended because, for some input dataset, their values fail to identify inconsistencies in the model (see chapter §4).
For each metric a validity interval has been defined (the best value could be a boundary of the interval or stay within the interval). Table 21 shows, for each metric, a three colour scale. If the value of the metric is outside the associated range, the square is coloured in bright red, otherwise the square is coloured from green to brown depending how close the metric is to the best value (green).

In `Table 24`, the groups of simulations are subdivided according to the different hydraulic models: in the table on the top the HVGM, in the centre the VGM and on the bottom the VGM-2cm. As mentioned earlier, the HVGM-2cm is excluded because the number of successful simulations that uses this hydraulic model is too restricted in order to allow a consistent analysis for the calibration programme.

Each group of inverse analysis is indicated with the identification code 1_1_xx (2nd column) and with a name including information on:

- the type of dataset used (wc, PH, wc+PH)
- the weighting distribution (w=1; w=TMA)
- the layer/s whose parameters are optimized in the simulation (1st column)

In `Table 24` all the 16 metrics compare simulated data to observed data over the whole simulation period (2018 year) and the values of the metrics are reported and coloured from green to red. If in a group of inverse analysis more than 6 metrics are outside the validity range, we consider that group unacceptable. The number of metrics that exceeds the valid range is reported in the 19th column (the bright red colour helps distinguish these metrics easily).

Appling the rules just explained, the following groups of inverse analysis could be excluded from the next calibration phase (1st phase of the calibration):

HVGM_embankment_wc_weight=TMA (1_1_36);
HVGM_foundation_PH_weight=TMA (1_1_41);
HVGM_embankment+foundation_PH_weight=TMA (1_1_45);
HVGM_embankment+foundation+fluvial_PH+wc, weight=TMA (1_1_47).
VGM_foundation_PH_w=TMA (1_1_42)
VGM_embankment+foundation_PH_w=TMA (1_1_46)
VGM_emb_w=TMA (1_1_37)

193 indirect simulations pass to the next calibration phase.

Looking closely at Table 24, other observations could be made:

- The hydraulic parameters of the embankment layer could be optimized in an effective way using HVGM or VGM both with a PH or w.c. observation dataset as input.
- The hydraulic parameters of the fluvial layer could be optimized in an effective way using HVGM or VGM with a w.c. observation dataset as input.

- The hydraulic parameters of the foundation layer could be optimized in an effective way using HVGM with a PH observation dataset as input.
- The inverse simulations that use the TMAT to allocate the weighs of the observation dataset perform worse with respect to the uniform distribution of weights (weights equal to unity).
- For the HVGM and VGM, NSE_j and NSE_{rel} have a relatively big dispersion of the values outside the accepted range or very close to the thresholds. The same could be observed for MBE, MAE, MSE (see Table 24).
- For the three hydraulic models (HVGM, VGM, VGM-2cm), R^2 has mostly values greater than 0,7, suggesting a good fit between observed-simulated dataset. The joint use of R^2 together with all the other metrics stresses the limits of the use of this parameter as single indicator often unable to catch the real performance of the simulation.
- A part from the seven groups of inverse analysis excluded, the others have a low number of metrics outside the valid range (low number of bright red cells), sign of a good quality of the simulations.
- The model performance evaluation metrics (AIC, BIC, HQIC) show a very low score for all 6 groups of inverse analysis that are going to be excluded. This demonstrates that model performance metrics could be used as unique set of indices in the preliminary phase to individuate the groups of simulations with a poor/insufficient performance.

HVGM		NSE	IA	NSE_{rel}	IA_{rel}	NSE_j	IA_j	MBE	MAE	IA'	CRM	MSE	KGE	AIC / AICmin	BIC / BICmin	HQIC / HQICmin	R^2	n° < threshold
Embankment_PH_weight=1	1_1_1	0,9	1,0	-3,8	1,0	0,8	0,9	-0,4	0,6	0,9	-0,3	0,7	0,7	6,0	6,0	6,0	1,0	1
Embankment_wc_weight=1	1_1_2	-0,2	0,8	-0,1	0,8	-0,2	0,6	0,0	0,0	0,6	0,0	0,0	0,4	0,4	0,4	0,4	0,7	0
Embankment_PH_weight=1	1_1_3	0,9	1,0	-3,8	1,0	0,8	0,9	-0,4	0,6	0,9	-0,3	0,7	0,7	6,0	6,0	6,0	1,0	1
Embankment_wc_weight=1	1_1_4	-0,5	0,8	-0,4	0,8	-0,3	0,6	0,0	0,0	0,6	0,0	0,0	0,3	0,3	0,3	0,3	0,7	0
Embankment_wc_weight=TMA	1_1_36	0,4	0,8	-0,7	0,8	0,4	0,7	-38,8	38,8	0,7	-0,2	47,0	0,5	20,8	20,7	20,8	0,8	6
Fluvial_wc_weight=1	1_1_7	0,8	0,8	0,7	0,8	0,4	0,7	0,0	0,0	0,7	0,0	0,0	0,7	0,2	0,2	0,2	0,8	0
Fluvial_wc_weight=TMA	1_1_39	-26,8	0,8	-26,6	0,8	-5,3	0,3	0,1	0,1	0,3	0,1	0,0	0,5	0,5	0,5	0,5	0,8	3
Foundation_PH_weight=1	1_1_9	0,6	0,8	-1,0	0,8	0,4	0,7	-0,4	0,6	0,7	-0,6	1,1	0,3	3,5	3,5	3,5	0,7	2
Foundation_PH_weight=TMA	1_1_41	0,0	0,4	0,7	0,4	0,1	0,5	-3,6	4,6	0,5	1,8	29,6	-1,0	14,0	14,0	14,0	0,5	9
Aquifer_PH_weight=1	1_1_17	0,3	0,9	0,3	0,9	0,2	0,7	-0,5	0,5	0,7	-0,1	0,5	0,7	1,0	1,0	1,0	0,8	0
Emb+found_PH_weight=1	1_1_13	0,8	0,9	-0,4	0,9	0,6	0,8	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,9	0
Emb+foun_PH_weight=TMA	1_1_45	-0,1	0,5	-0,1	0,5	0,2	0,5	-12,5	13,4	0,5	1,0	122,7	-0,3	23,2	23,1	23,1	0,8	7
Emb+found+fluv_PH+wc_w=1	1_1_15	0,8	0,9	0,9	0,9	0,7	0,8	-0,1	0,3	0,8	-1,2	0,3	-0,3	4,4	4,4	4,4	0,9	1
Emb+found+fluv_PH+wc_TMA	1_1_47	-0,1	0,5	-0,2	0,5	0,2	0,5	-12,9	13,6	0,5	1,0	122,1	-0,3	11,6	11,6	11,6	0,8	7
Emb+fluv_wc_weight=1	1_1_11	-0,1	0,8	-0,1	0,8	0,0	0,6	0,0	0,0	0,6	0,0	0,0	0,3	0,6	0,6	0,6	0,7	0
VGM																		
Embankment_PH_weight=1	1_1_5	0,9	1,0	-3,9	1,0	0,8	0,9	-0,4	0,6	0,9	-0,3	0,7	0,7	5,7	5,7	5,7	1,0	1
Embankment_wc_weight=1	1_1_6	-0,1	0,8	-0,1	0,8	0,0	0,6	0,0	0,0	0,6	0,0	0,0	0,3	0,6	0,6	0,6	0,7	0
Fluvial_wc_weight=1	1_1_8	0,8	0,8	0,7	0,8	0,4	0,7	0,0	0,0	0,7	0,0	0,0	0,0	0,3	0,3	0,3	0,8	0
Foundation_PH_weight=1	1_1_10	0,6	0,8	-1,0	0,8	0,4	0,7	-0,6	0,7	0,7	-0,6	1,1	0,3	3,3	3,3	3,3	0,7	4
Foundation_PH_weight=TMA	1_1_42	0,0	0,4	0,7	0,4	0,1	0,5	-3,6	4,6	0,5	1,8	29,6	-1,0	13,2	13,2	13,2	0,5	9
Aquifer_PH_weight=1	1_1_18	0,3	0,8	0,3	0,8	0,2	0,6	-0,5	0,5	0,6	-0,1	0,5	0,7	1,0	1,0	1,0	0,7	2
Emb+found_PH_weight=1	1_1_14	0,8	0,9	-0,4	0,9	0,6	0,8	-0,2	0,5	0,8	0,6	0,6	0,3	5,6	5,6	5,6	0,9	0
Emb+found_PH_w=TMA	1_1_46	-0,1	0,5	0,1	0,5	0,2	0,5	-11,8	12,8	0,5	1,0	121,9	-0,3	21,8	21,8	21,8	0,7	7
Emb+found+fluv_PH+wc_w=1	1_1_16	0,8	0,9	0,9	0,9	0,7	0,8	-0,1	0,3	0,8	-1,2	0,3	-0,3	4,2	4,2	4,2	0,9	1
Emb+fluv_wc_w=1	1_1_12	0,9	0,9	0,8	0,9	0,7	0,9	0,0	0,0	0,9	0,0	0,0	0,8	0,8	0,8	0,8	0,9	0
Fluvial_wc_weight=TMA	1_1_40	-39,5	0,7	-39,2	0,7	-8,1	0,1	0,1	0,1	0,1	0,2	0,0	0,4	0,5	0,5	0,5	0,7	3
Emb_PH_w=TMA	1_1_37	0,3	0,9	0,5	0,9	0,4	0,7	-39,6	39,6	0,7	-0,2	55,5	0,4	20,1	20,1	20,1	0,9	6
Emb+fluv_wc_w=TMA	1_1_44	0,4	0,7	0,1	0,7	0,5	0,7	0,0	0,0	0,7	0,0	0,0	0,5	0,9	0,9	0,9	0,5	1
VGM-2cm																		
Embankment_wc_weight=1	1_1_20	-0,4	0,9	-0,4	0,9	-0,3	0,6	0,0	0,0	0,6	0,0	0,0	0,0	0,3	0,3	0,3	0,7	0
Fluvial_wc_weight=TMA	1_1_24	0,8	0,9	0,8	0,9	0,5	0,7	0,0	0,0	0,7	0,0	0,0	0,7	1,0	1,0	1,0	0,8	0
Emb+fluv_PH_weight=1	1_1_28	-0,7	0,7	-2,0	0,7	-0,1	0,5	-0,3	0,5	0,5	-0,5	0,7	0,4	2,2	2,2	2,2	0,7	2
Aquifer_PH_weight=1	1_1_34	0,3	0,8	0,3	0,8	0,3	0,7	-0,5	0,5	0,7	-0,1	0,5	0,6	1,0	1,0	1,0	0,7	0
VGM+2cm_foundation_PH_w=1	1_1_26	-0,7	0,7	-2,0	0,7	-0,1	0,5	-0,3	0,5	0,5	-0,5	0,7	0,4	2,2	2,2	2,2	0,7	2
BEST FIT		1	1	1	1	1	1	0	0	1	0	0	1	1	1	1	1	

Table 24: Set of metrics applied to the different groups of inverse analysis in order to investigate the behaviour during the whole 2018 simulation period. The last column reports the number of metrics that does not comply the minimum threshold set for each metric. In red the groups of inverse analysis that have shown a poor performance, below the require threshold. It is important to report that the following groups of inverse analysis have a n° of indirect simulation <3 within the group: 1_1_47, 1_1_11, 1_1_12, 1_1_44, 1_1_46, 1_1_20, 1_1_26 (2 simulations); 1_1_37, 1_1_40 and 1_1_28 (1 simulation).

7_2_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

A preliminary evaluation of the available 33 groups of inverse analysis has been carried out. A set of metrics (accuracy, efficiency and model selection criteria) has been chosen and for each of them a threshold is set. Metrics/indices are computed for each simulation on the whole simulation period in order to compare observed-simulated datasets then an average of the values assumed by each metric between the indirect simulations belonging to the same group is computed. This is legitimate because the simulations that belong to the same group optimize the same layer, using the same hydraulic model and the same weighting distribution. In this preliminary phase it is interesting to look at the global behavior of the groups of indirect simulations in order to exclude from the following phases (which look at the single indirect simulations) the ones that behave poorly. Moreover the modeler has an initial general idea of the performance of the groups of simulations and this helps to set the score scale and the threshold values for the following phases in which each indirect simulation is investigated with more accuracy. In this preliminary evaluation, metrics/indices that in chapter 4 ("preliminary tests on the metrics/indices performance") showed in some cases inconsistent behaviour have been used. This is because they are able to give a general indication on the performance of groups of indirect simulations but they will be excluded from the following calibration steps because not able to evaluate singularly the indirect simulations making hard to compare their performance and grading them on a scale (from the best to the worst). N° 16 metrics/indices have been computed for each group of inverse analysis and 6/16 metrics values outside the valid range determine the exclusion of the group for an insufficient performance. Seven groups of inverse analysis out of 33 are excluded according to the selection just exposed. N° 193 indirect simulations pass to the first calibration phase (chapter 7_3). It has been observed that a part from the seven groups of inverse analysis excluded, the remaining groups have a good/very good performance with a low number of metrics/indices below the accepted thresholds. Moreover it has been observed that the Model Selection Criteria Metrics (AIC, BIC HQIC) show an insufficient score for all the seven excluded groups of inverse analysis while for the remaining groups a score always above the set threshold. This means that these metrics are a good indicator for a preliminary evaluation of the performance of groups of inverse analysis. It has been observed also that a weighting distribution TMAT shows in all the cases a poorer performance of the indirect simulations with respect to a weighting distribution with all the data equal to unity (w=1).

In the following subchapter 7_3 the first calibration phase is applied to each indirect simulation: a set of chosen metrics/indices is used to quantify the discrepancies between observed and simulated datasets. While the pre-calibration phase is applied to groups of indirect simulations, the first phase is applied to the single indirect simulation in order to allow a direct comparison between them with the final purpose of individuate the best set/s of optimized parameters.

7_3 PERFORMANCE EVALUATION OF INDIRECT SIMULATIONS: THE FIRST PHASE OF THE CALIBRATION PROCESS

The preliminary phase, presented in chapter §7_2, to evaluate the performance of groups of inverse analysis could be useful when a large number of indirect simulations is available in order to discard easily the ones that are below the expected performance and focus the attention on a restricted number of simulations. The advantage of performing this preliminary phase stays in the more efficient way of investigating the remaining simulations, saving time and having a first general overview on the limits and potentialities of the available set of indirect simulations.

The new phase of the calibration process that is going to be presented, could be applied at the end of the previous step to the purged pool of simulations or, in case of a small number of inverse analysis, to the whole pool of indirect simulations (skipping the preliminary phase). Differently from the preliminary phase (chapter §7_2), the first phase of the calibration investigates the performance of each indirect simulation and not the performance of the groups of inverse analysis.

Indirect simulations performance is examined using statistic metrics. A scoring-based methodology is developed and applied to the case study. The scores are given according to the ability of each metric to predict the performance of the simulation and according to the distance of its value from the best fit. In this first phase of the calibration process, the attention is focused on the simulation output that is compared to the observation data; in other words the differences between observed-simulated dataset is under scrutiny.
Metrics are used to interpret the performance of each indirect simulation over the whole temporal period (November 2017 to June 2018) and over the 12 flood peak-periods (from A to N) and 12 off-peak periods (from A' to O'). The peak- periods and off-peak periods have been presented in chapter §6_1 and in Figure 19. The investigation of the behaviour of the indirect simulations over the flood peak periods is of primary importance in a calibration phase in view of future stability analysis which need reliable pwp distributions. It is extremely important to avoid an evaluation of the performance of the simulation based only on the behaviour over the whole simulation period because it is not indicative of a good representation of the peaks (flood periods) but often only of the off-peak periods. The set of metrics/indices used in the first phase of the calibration could be used similarly for the evaluation of the performance of direct simulations when a reliable observation dataset is available.

The indices used to investigate the behaviour of the indirect simulations over the **''whole simulation period**'' are the following:

- **accuracy metrics**: the coefficient of determination R^2 and r^2, the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Modified Root Mean Square Error (MRMSE) that considers the weight associated to each measurement of the input dataset and the number of estimated parameters,

- -**efficiency metrics**: the Nash-Sutcliffe Efficiency Criteria (NSE), the modified form of NSE (NSE_j), the Relative Efficiency Criteria (NSE_{rel}), the modified Index of Agreement (IA'), the Kling-Gupta efficiency (KGE),

- -**Model selection criteria**: the Akaike Information Criteria (AIC), the Bayesian Information Criteria (BIC) and the Hannan Information Criteria (HQIC).

The use of metrics has been associated to the use of another parameter which takes into account the input data (in terms of water content, pressure head or both). It is preferable to consider approximately the same number of efficiency and accuracy metrics in order to give a ''balanced'' evaluation of the performance of an indirect simulation.

The metrics used to investigate the flood **''peak periods''** and **''off-peak periods''** are the following:

- **accuracy metrics**: the coefficients of determination (R^2 and r^2), MAE, RMSE.
- **efficiency metrics**: IA', NSE, NSE_j, NSE_{rel}
- **statistical significance criteria**: the logical t-test (to assess if the difference between simulated-observed values is statistically significant or not).

As could be observed, a smaller number of metrics is applied to single time periods in order to avoid the confusion given by a high number of metrics to process for 24 different time periods (12 peak periods and 12 off-peak periods).

The choice of the metrics to use fell on MAE and RMSE that are among the ''best'' overall measures of a model performance, as they summarize in an easy way the mean difference between observed and predicted data (Willmott, 1982). R^2 has been included in the pool of metrics because it is worldwide used in model evaluation and even though, in some cases, significant values (close to 1) of R^2 could be misleading because they do not correspond to equally good simulations, in the majority of the cases R^2 is able to catch the overall behaviour of the simulation and communicate this information in a very understandable way. Its use together with other accuracy metrics, efficiency metrics and the significance test ensure to catch possible misleading evaluations. Model selection criteria are not used in the evaluation of the performance during the peak/off-peak periods because, as reported in chapter §2_7_2_5, these indices are suitable for a comparison between models but not for comparisons ''inside'' the same simulation.

IA, IA_{rel} and IA_j have been excluded from the chosen indices because, in some cases, an inconsistent behaviour has been shown for some input datasets (see chapter §4). Over these metrics, the modified form IA' is preferred. The trilogy composed of NSE and its modified forms (NSE_j and NSE_{rel}) are used because each of them has particular advantages: the modified form NSE_j is less sensitive to overestimation of large errors while the relative form NSE_{rel} avoids an over or under prediction of the higher values and a low influence of the lower values resulting in a better overall estimation. So globally the efficiency indices of the first phase of the calibration are NSE, NSE_j, NSE_{rel} and IA'. For what concerns the accuracy criteria only two metrics have been chosen: RMSE and MAE. MAE is chosen over ME (MBE) because errors with different signs cancel out in ME, as previously explained, while RMSE is preferred over MSE because more known and used in simulation evaluation but substantially these two indices are equivalent (RMSE is the square root of MSE). MAE is undoubtedly a better metric compared to RMSE but RMSE has the benefit of penalizing large errors (due to errors that are squared before being averaged), that, in our case, is extremely appropriate. Both the forms of RMSE are considered: the original one and the one that takes into consideration the weights of the input dataset and the degree of freedom (for more information see chapter §2_7_2_1). Due to the higher information content in the modified form of RMSE, a greater score is given to this parameter with respect to the original form (see Table 26). For what concerns the coefficients of determination, the two forms R^2 and r^2 are equivalent when the weights of the dataset are all equal to 1, so only for the dataset with weights according to TMA both the forms are used.

Despite some inconsistencies observed in chapter §4, the KGE metric is used in the first phase of the calibration process. In fact as demonstrated by Gupta et al (2009) the KGE metric represents an improvement of the widely used NSE criterion because it considers three different types of model errors namely the variability bias, mean bias and correlation resolving several perceived shortcomings in NSE. In Table 25 the scores associated to the different values of each metric are presented. As it is possible to observe score reduces as the value of the metric moves away from the best fit.

R^2		**RMSE**		**MAE**		**IA'**	
$R^2 \leq 0{,}50$	0	RMSE≥1	0	MAE≥0,60	0	IA'≤0	0
$0{,}50 < R^2 < 0{,}75$	0,50	0,50<RMSE<1	0,25	0,30<MAE<0,60	0,5	0<IA'<0,50	0,25
$0{,}75 < R^2 \leq 1$	1	0,25≤RMSE<0,50	0,50	MAE≤0,30	1	0,50<IA'<0,75	0,50
		RMSE<0,25	1			IA'≥0,75	1
AIC		**BIC**		**HQIC**		**TYPE OF DATASET**	
AIC≥10	0	BIC≥10	0	HQIC≥10	0	PRESS.HEAD	1
7<AIC<10	0,25	7<BIC<10	0,25	7<HQIC<10	0,25	WATER CONT.	0,50
4≤AIC≤7	0,50	4≤BIC≤7	0,50	4≤HQIC<7	0,50	PH+wc	0,75
AIC<4	1	BIC<4	1	HQIC<4	1		

KGE		NSE		NSE_{rel}		NSEj (j=1)	
KGE≥0,7	1	NSE≤0	0	NSE_{rel}≤0	0	NSEj≤0	0
0,5<KGE<0,7	0,5	0<NSE<0,5	0,25	0<NSE_{rel}<0,5	0,25	0<NSEj<0,5	0,25
0,25<KGE≤0,50	0,25	0,5≤NSE≤0,75	0,50	0,50≤NSE_{rel}≤0,75	0,50	0,50≤NSEj≤0,75	0,50
0≤KGE≤0,25	0	0,75<NSE≤1	1	0,75<NSE_{rel}≤1	1	0,75<NSEj≤1	1

Table 25: Scores associates to the different values of each metric used in phase 1 of the calibration programme

In this first calibration phase each indirect simulation is examined (193 for the case study). The goal is to assign a significant score to each indirect simulation in order to draw up a final ranking that highlights the best performing simulations.
To do so the type of input data and the accuracy of the simulation in describing flood and off-flood events are taken into consideration. In the definition of the final score metrics/indices and a system of weights attributed to the single metric and to the different groups of metrics are used.
Let's proceed step by step to define our scoring strategy for the ''whole simulation period'', for the ''peak periods'' and ''off-peak periods''.

For each indirect simulation:

EVALUATION OF THE PERFORMANCE ON THE WHOLE SIMULATION PERIOD

- We evaluate the metrics which measure the distance between the model output and the observed data on the whole simulation period (R^2, r^2, MAE, RMSE, MRMSE, NSE, NSE_j, NSE_{rel}, IA', KGE, AIC, BIC, HQIC).
- Table 22 defines how to assign a score to each metric depending on the closeness of its value to the best fit.
- All the considered metrics are categorized into 8 major groups: see table 23a, first column.
- In Table 23a second column, the weight (ω_{int}) of each metric belonging to one of the 8 groups is set.
- In Table 23a third column, the single weight (ω_{ext}) attributed to each group of metrics is set.
- In Table 23a fourth column, the final weight (ω_{fin}) attributed to each metric of the 8 groups is set.

$$\omega_{fin} = \omega_{ext}\omega_{int} \quad \text{[equation 131]}$$

The indirect simulation achieves the maximum score (5,5) if all the metrics obtain the best score (1) in accordance with `Table 25`.
Let's make an example considering only the metrics of group 4 (NSE, NSE_j, NSE_{rel}).
Assuming that NSE=0,85; NSE_j=0,70; NSE_{rel}=0,78, according to `Table 25` the scores are 1, 0,5 and 1 respectively.
The value of ω_{int} is equal to 0,33 for the three metrics of group 4
The value of ω_{ext} of group 4 is equal to 0,5.
The value of ω_{fin} is equal to 0,5*0,33=0,17 for each metric of group 4
The score of group 4 (s_{group4}) and the final score of the indirect simulation (s_{fin}) are the following in the considered case:

$$s_{group4} = 1 * 0{,}17 + 0{,}70 * 0{,}17 + 1 * 0{,}17 = 0{,}425$$

$$s_{fin} = \sum_{i=1}^{8} s_{group_i} \quad \text{[equation 132]}$$

Where i is the number of groups of metrics.

EVALUATION OF THE PERFORMANCE ON THE PEAK PERIODS

Now the indirect simulation performance is investigated in the 12 flood periods, following a similar approach.

- We evaluate the metrics which measure the distance between the model and the data at a certain flood event (R^2, r^2, MAE, RMSE, NSE, NSE_j, NSE_{rel}, t-test).
- Table 22 defines how to assign a score to each metric depending on the closeness of its value to the best fit.
- All the considered metrics are categorized into 6 major groups: see table 23b, first column.
- In Table 23b second column, the weight (ω_{int}) of each metric belonging to one of the 8 groups is set.
- In Table 23b third column, the single weight (ω_{ext}) attributed to each group of metrics is set.
- In Table 23b fourth column, the final weight (ω_{fin}) attributed to each metric belonging to one of the 8 groups is set.
- We repeat the procedure for each flood period. At the end of the procedure we have 12 scores, one for each flood period. Table 25 gives the partition coefficients c_i (weights attributed to each flood period) and the final score of the considered indirect simulation (s_{fin}) is computed as follows:

$$s_{fin} = \sum_{j=1}^{12} c_j \sum_{i=1}^{6} s_{group_i} \qquad \text{[equation 133]}$$

Where j are the number of flood periods (12); c_j is the partition coefficient, i is the number of groups of metrics (6).

The indirect simulation achieves the maximum score (7,13) if all the metrics obtain the best score (1) in accordance with `Table 25`.

EVALUATION OF THE PERFORMANCE ON THE OFF-PEAK PERIODS

Now the indirect simulation performance is investigated at the 12 off-peak periods, following the same approach specified for the peak periods.

The indirect simulation achieves the maximum score (2,38) if all the metrics obtain the best score (1) in accordance with `Table 25`.

Consequently an indirect simulation can achieve the maximum score of 15 (5,5+7,13+2,38).

A greater score is given to the flood peaks periods (7,13) compared to the score attributed to the drops of the river water level (2,38) and to the whole simulation period (5,5) because proposed methodology targets toward the final aim of the calibration which is the elaboration of a reliable pwp distribution in the most significant instants of the simulation.

For a bank section in fine-grained materials, as the one in the presented case study, the most significant instants of the simulation are the floods more persistent in time that reached medium-high water levels. Off less interest for future stability analysis, the performance of the indirect simulations in the off-peak periods. Looking at the whole simulation period can be interesting because it could give an overall idea on the simulation performance without going into details.

The score given to the flood peak-periods (7,13) is subdivided among the different peaks according to the water height reached and the persistence of the flood in time. The area of each flood has been measured then divided for the sum of the areas of the 12 floods that characterize the simulation period.

The partition coefficient for each flood is equal to [see equation 134 and 135]:

$$c_i = \frac{\mathrm{Ai}}{\sum_{i=1}^{n} \mathrm{A_i}}$$ [equation 134]

$$\sum_{i=1}^{n} c_i = 1$$ [equation 135]

where A_i is the area underneath the curve in the graph time -water height (m a.s.l), n is the number of flood-periods (12). In Figure 24 the areas of nine (over 12) flood-periods (named from ''A'' to ''I'', see Figure 20) are presented.

The procedure adopted to evaluate the partition coefficients for the off-peak periods is the same: partition coefficients are attributed according to the persistence of the period and the water level reached (floods of minor entity and relevance with respect to the 12 peaks are observed in these periods).
To perform the inverse analysis, the whole set of available monitoring points has been used in the objective function in order to test the performance of the simulation in the prediction of the pwp and water content distributions in the different layers where sensors are available.

In Table 26 a summary of the weights given to the chosen metrics is reported. The table is divided in the weights given to the single metrics of each group (''internal weight'') in the first column, in the weight given to each group with respect to the other groups based on their relevance in the performance evaluation (''external weight'') in the second column, and the final weight obtained as the product of the internal weight for the external weight for each metric in the third column. The ''external weights'' are responsible of an equal subdivision of the weights between the group of metrics (efficiency metrics, accuracy metrics and model selection criteria) in order to have a final score for each simulation which is the equilibrated result of the score obtained for each type of metric. The sum of the ''final weights'' of each metric is equal to the maximum score that a simulation could have (5,50 points for the ''whole period'', 7,13 points for the ''peak -periods'', 2,38 for the ''off-peak periods'').

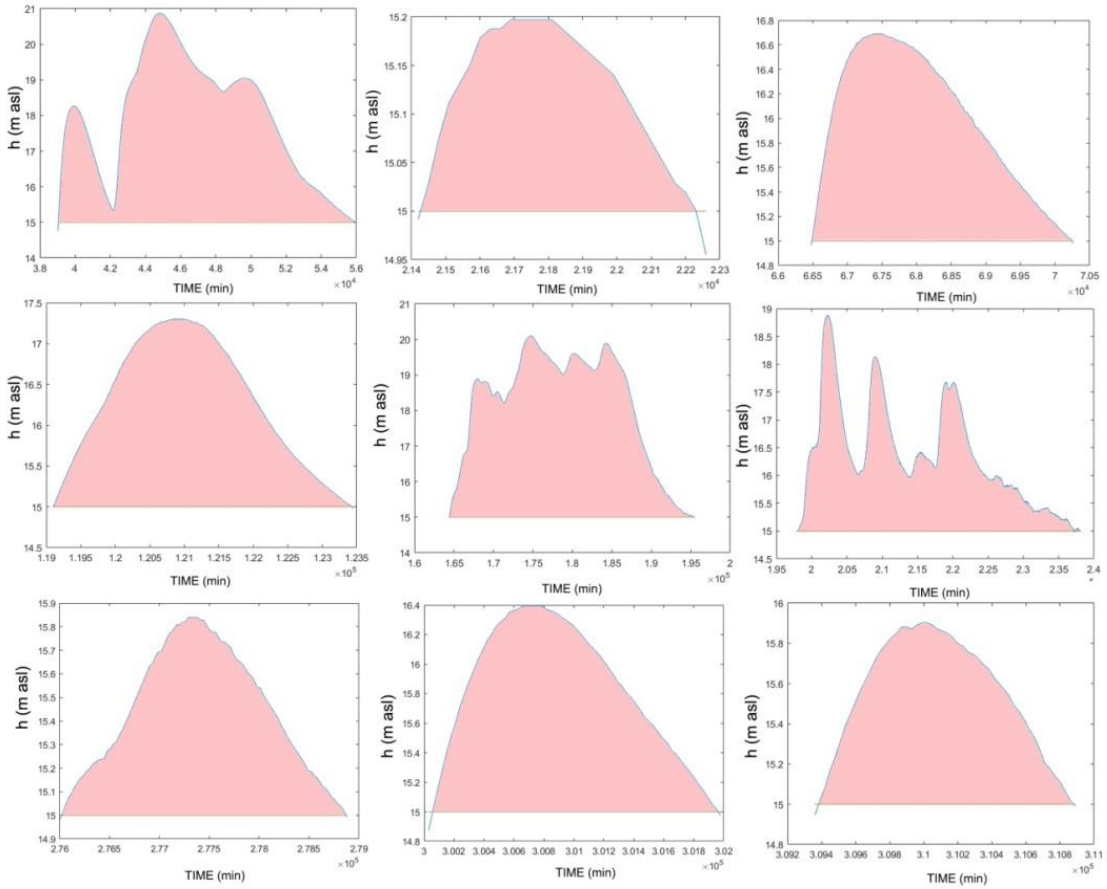


Figure 24: Area of each flood under the curve water height (m a.s.l.)-time (hydrograph)

In Table 27 the partition coefficients c_i are reported for each pick-period. The points P_i for each period are calculated as follows:
If S_i is the score related to the i_th flood, the final score related to the indirect simulation is:

$$S = \sum_{i=1}^{12} c_i \cdot S_i \qquad \text{[equation 136]}$$

The maximum score for the i-th peak-period:

$$\max\_S_i = c_i \cdot 7{,}13$$
$$\sum_{i=1}^{n} \max\_S_i = 7{,}13 \qquad \text{[equation 137]}$$

Similarly the maximum score for the i-th off-peak period is:

$$\max\_S_i = c_i \cdot 2{,}38$$
$$\sum_{i=1}^{n} \max\_S_i = 2{,}38 \qquad \text{[equation 138]}$$

For each simulation, the points from the evaluation of the performance on the ''whole simulation period'' and on the ''peak and off-peak periods'' are summed.

The simulations whose final score is greater than 7,5/15 are considered acceptable and they will undergo the second phase of the calibration programme.

It is relevant to notice that in the preliminary phase (chapter §7_2) no scores have been given based on the goodness of the values of the metrics. It has been chosen only a range of acceptance (e.g. for the coefficient of determination R^2 between 0,5-1 acceptable, between 0-0,5 not acceptable) while for the first phase of the calibration the score is in some way proportional to the goodness of the metrics. The main reason is that the first phase is a refined method of investigation of the set of available inverse analysis. The main goal is the identification of the most promising simulations which will pass to the next calibration phase while the preliminary phase is intended as a rapid initial analysis, unable to hierarchize the set of simulations but select only the groups of simulations that reach a minimum performance.

Table 27a

GROUP OF METRICS APPLIED TO THE WHOLE SIMULATION PERIOD	INTERNAL WEIGHT OF EACH METRIC	EXTERNAL WEIGHT (WITH RESPECT TO THE OTHER GROUPS OF METRICS)	FINAL WEIGHT OF EACH METRIC
R^2 r^2	0,5 0,5	1	0,5 0,5
RMSE $RMSE_{modif}$	0,4 0,6	0,5	0,2 0,3
MAE	1	0,5	0,5
NSE NSE_{rel} NSE_j	0,3 0,3 0,3	0,5	0,17 0,17 0,17
IA'	1	0,5	0,50
Type of measurement	1 for PH, 0,5 for wc, 0,75 for PH+wc	1	1
AIC BIC HQIC	0,3 0,3 0,3	1	0,33 0,33 0,33
KGE	1	0,5	0,5
sum			5,50

Table 27b			
GROUP OF METRICS APPLIED TO THE PEAK PERIODS	INTERNAL WEIGHT OF EACH METRIC	EXTERNAL WEIGHT (WITH RESPECT TO THE OTHER GROUPS OF METRICS)	FINAL WEIGHT OF EACH METRIC
R^2 r^2	0,5 0,5	1,43	0,71 0,71
MAE	1	1,43	1,43
RMSE	1	1,43	1,43
NSE NSE_{rel} NSE_j	0,3 0,3 0,3	0,71	0,24 0,24 0,24
IA'	1	0,71	0,71
h_{logic}	1	1,43	1,43
sum			7,13

Table 27c			
GROUP OF METRICS APPLIED TO THE OFF-PEAK PERIODS	INTERNAL WEIGHT OF EACH METRIC	EXTERNAL WEIGHT (WITH RESPECT TO THE OTHER GROUPS OF METRICS)	FINAL WEIGHT OF EACH METRIC
R^2 r^2	0,5 0,5	0,48	0,24 0,24
MAE	1	0,48	0,48
RMSE	1	0,48	0,48
NSE NSE_{rel} NSE_j	0,3 0,3 0,3	0,24	0,08 0,08 0,08
IA'	1	0,24	0,24
h_{logic}	1	0,48	0,48
sum			2,38

Table 26a,b,c: Summary of the weights attributed to the metrics for the evaluation of the performance on the ''whole simulation period'' (first table), ''peak-periods'' (second table) and ''off-peak period''(third table). The internal weight (with respect to the other metrics of the group), external weights (with respect to the other groups of metrics) and the final weight (attributed to each metric) are specified.

	A	B	C	D	E	F	G	H	I	L	M	N	SUM
PARTITION COEFF FOR EACH PEAK PERIOD	0,003	0,0004	0,213	0,016	0,024	0,006	0,419	0,235	0,067	0,005	0,007	0,004	
MAX SCORE FOR EACH PEAK PERIOD	0,023	0,003	1,518	0,115	0,171	0,041	2,985	1,675	0,480	0,039	0,050	0,026	**7,13**
	A'	B'	C'	D'	E'	F'	G'	H'	I'	L'	M'	N'	SUM
PARTITION COEFF FOR EACH OFF-PEAK PERIOD	0,14	0,12	0,05	0,25	0,06	0,03	0,03	0,01	0,04	0,11	0,04	0,11	
MAX SCORE FOR EACH OFF-PEAK PERIOD	0,33	0,29	0,12	0,60	0,14	0,08	0,08	0,04	0,09	0,26	0,09	0,26	**2,38**

Table 27:The table shows the partition coefficients , which allow a subdivision of the points in the different flood peak-periods and off- peak-periods.

In Table 28 the statistics of the first phase of the calibration are presented. The initial number of simulations is 193, the number of acceptable simulations (with score >7,5/15) is 155 (80% of the initial number of simulations). The 60,6% of the acceptable simulations have a score between 7-9; the 20% between 9-11, 18,1% between 11-13 and 1,3% between 13-15.

Max score	15
Threshold value	7,5
Initial n° of simulations	193
n° of accepted simulations	155
n° of excluded simulations	38
n° of simulations with score between 7-9	94
n° of simulations with score between 9-11	31
n° of simulations with score between 11-13	28
n° of simulations with score between 13-15	2

Table 28: Statistics of the indirect simulations tested in the first phase of the calibration process.

To perform an easy and rapid comparison between the results of the preliminary phase and the first phase of the calibration, the groups of inverse analysis could be considered instead of an analysis simulation by simulation (very difficult due to their high number). As reported in chapter §7_2, considering the average value of a metric as representative of a group of inverse analysis is legitimate because the interval of variance of these values is extremely limited within the same group (as could be observed also in Table 33) and for this reason the average value is a good indicator of the performance of a group.
Table 29 presents the number of successful simulations in each group of inverse analysis that have overcome the first calibration phase (third column) and the percentage of successful simulations in each group with respect to the initial number (fourth column). The fifth column presents the hydraulic parameters optimized in each group of inverse analysis. As could be observed in Table 29:

- -the higher percentage of successful simulations (that overcome the first calibration phase) is observed for the VGM (87%) and the lower for the VGM-2cm (26%).
- -the higher percentage of successful simulations is observed for the HVGM using a mixed dataset of observation points (wc+PH) (71%) and for the VGM using a PH dataset (81%). The performances of the indirect simulations using a dataset of type ''water content'' are poorer compared to the same simulations (that optimise the same set of parameters and using the same weighting distribution) that use a PH or wc+PH dataset.
- -The use of the weighting distribution TMA assures the lower percentage of successful simulations (23% for the VGM; 17% for the HVGM).
- -The VGM-2cm shows the lower percentage of successful simulations (which means a poor performance of the simulations) using both the weighting distribution w=1 (27%)and w=TMA (25%).

GROUPS OF INVERSE ANALYSIS	IDENT. CODE	N° OF SUCCESSFUL SIMULATIONS	% SUCCESSFUL SIMULATIONS	OPTIMIZED PARAMETERS
HVGM_emb_PH_w=1	1_1_1	13	100	$\theta s,\theta r,k,n,\alpha,k_w,\alpha_w,\theta_m$
HVGM_emb_wc_w=1	1_1_2	8	100	$\theta s,\theta r,k,n,\alpha,k_w,\alpha_w,\theta_m$
HVGM_emb_PH_w=1	1_1_3	14	100	$\theta s,\theta r,k,n,\alpha,\alpha_{w,\theta m}$
HVGM_emb_wc_w=1	1_1_4	8	100	$\theta s,\theta r,k,n,\alpha,\alpha_w$
HVGM_emb_PH_w=TMA	1_1_36	0	0	-
HVGM_fluv_wc_w=1	1_1_7	5	100	$\theta s,\theta r,k,\alpha$
HVGM_fluv_wc_w=TMA	1_1_39	3	100	$\theta s,\theta r,k$
HVGM_found_PH_w=1	1_1_9	9	100	$\theta s,\theta r,k,n,\alpha$
HVGM_found_PH_w=TMA	1_1_41	0	0	-
HVGM_aquifer_PH_w=1	1_1_17	9	100	$\theta s,\theta r,k,n,\alpha,k_w$
HVGM_emb+found_PH_w=1	1_1_13	8	100	$\theta s,\theta r,k,\alpha$
HVGM_emb+found_PH_w=TMA	1_1_45	0	0	-
HVGM_emb+found+fluv_PH+wc_w=1	1_1_15	5	100	$\theta r,k,\alpha,\alpha_w$
HVGM_emb+found+fluv_PH+wc_w=TMA	1_1_47	0	0	-
HVGM_emb+fluv_wc_w=1	1_1_11	2	100	$\theta s,\alpha$
VGM_emb_PH_w=1	1_1_5	12	100	$\theta s,\theta r,k,n,\alpha$
VGM_emb_WC_w=1	1_1_6	5	100	$\theta s,k,n,\alpha$
VGM_fluv_w.c._w=1	1_1_8	4	100	$\theta s,k,\alpha$
VGM_found_PH_w=1	1_1_10	12	100	$\theta s,\theta r,k,n,\alpha$
VGM_found_PH_w=TMA	1_1_42	0	0	-
VGM_aquifer_PH_w=1	1_1_18	13	100	$\theta s,\theta r,k,n,\alpha$
VGM_emb+found_PH_w=1	1_1_14	7	100	$\theta s,\theta r,k,n,\alpha$
VGM_emb+found_PH_w=TMA	1_1_46	0	0	-
VGM_emb+found+fluv_PH+wc_w=1	1_1_16	8	100	$\theta s,\theta r,k,n,\alpha$
VGM_emb+fluv_wc_w=1	1_1_12	1	100	θs
VGM_emb_PH_w=TMA	1_1_37	0	0	-
VGM_emb+fluv_wc_w=TMA	1_1_44	2	100	$\theta s,k,n,\alpha$
VGM_fluv_wc_w=TMA	1_1_40	1	100	k
VGM+2cm_emb_wc_w=1	1_1_20	2	100	n,α
VGM+2cm_fluvial_WC_weight=TMA	1_1_56	3	25	$\theta s,k,n,\alpha,\theta r$
VGM+2cm_emb+fluv_PH_weight=1	1_1_28	1	100	k,n
VGM+2cm_aquifer_PH_weight=1	1_1_34	0	0	-

% of accepted simulations_VGM	0,87	% of accepted simul_VGM_w=1	1
% of accepted simulations_HVGM	0,85	% of accepted simul_VGM_w=TMA	0,23
% of accepted simulations_VGM+2cm	0,26	% of accepted simul_HVGM_w=1	1
% of accepted simulations_HVGM_PH	0,80	% of accepted simul_HVGM_w=TMA	0,17
% of accepted simulations_HVGM_wc	1,00	% of accepted sim_VGM+2cm_w=1	0,27
% of accepted simulations_HVGM_wc+PH	0,71	% of accep. sim_VGM+2cm_w=TMA	0,25
% of accep. simulations_VGM_PH	0,81		
% of accep.simulations_VGM_wc	1,00		
% of accep simul._VGM_wc+PH	1,00		

Table 29: The table presents the n° of indirect simulations in each group of inverse analysis that have overcome the first calibration phase (third column) and the % of these simulations with respect to the initial number of simulations in the same group (fourth column). In the last column the hydraulic parameters that are optimized in each group (we have to remember that for the groups of inverse analysis that optimize more than one layer at a time, a set of optimized parameters is obtained for each layer). The groups of inverse simulations that present zero successful simulations (performance below the imposed threshold) will be excluded from the second calibration phase as 1_1_36; 1_1_41; 1_1_45; 1_1_47; 1_1_42; 1_1_46; 1_1_37 and 1_1_34.

Table 30 presents the percentages of successful simulation for hydraulic model (HVGM, VGM, VGM-2cm) and layer whose parameters are optimized (embankment, fluvial, foundation, aquifer, embankment+fluvial, embankment+foundation, embankment+foundation+fluvial). As it is possible to observe:

- the HVGM has difficulties in the optimization of the hydraulic parameters of the foundation layer (60% of successful simulations after applying the first phase of the calibration) and in the optimization of groups of layers (73% for embankment+foundation and 71% embankment+foundation+fluvial). Almost the same percentages are observed for the VGM. The difficulties encountered in the optimization of the foundation layer could be linked to the lower number of observation points/sensors located in this layer (3 sensors monitoring the water potential) with respect to the other layers. For the same reason the optimization of groups of layers that include the foundation shows a poor performance (comparison between embankment+fluvial and embankment+foundation)
- No successful simulations optimizing 2 or more layers are available for the HVGM-2cm. Unlike what has been found for VGM and HVGM, the optimization of the hydraulic parameters of the fluvial layer shows the greatest difficulties for VGM-2cm (although it must be emphasized the lower initial number of indirect simulations available for this hydraulic model).

	HVGM	VGM	VGM-2cm
n° of indirect simulations for the optimization of the embankment layer	43 (91%)	17 (94%)	2 (100%)
n° of indirect simulations for the optimization of the fluvial layer	8 (100%)	5 (100%)	3 (25%)
n° of indirect simulations for the optimization of the foundation layer	9 (60%)	12 (67%)	1 (100%)
n° of indirect simulations for the optimization of the aquifer layer	9 (100%)	13 (100%)	0
n° of indirect simulations for the optimization of the embankment+fluvial layers	2 (100%)	3 (100%)	0
n° of indirect simulations for the optimization of the embankment+foundation layers	8 (73%)	7 (70%)	0
n° of indirect simulations for the optimization of the embankment+foundation+fluvial layers	5 (71%)	8 (100%)	0
sum	84	65	6
total n° of successful simulations that have overcome the 1st phase	**155**		
%successful simulations with respect to the initial number	**80%**		

Table 30: Number of simulations that have overcome the first phase of the calibration subdivided for layer and hydraulic model (HVGM, VGM, VGM-2cm). In round brackets the percentage of simulations that have passed the second phase with respect to the initial number of simulations falling into that category.

In `Table 31` the scores of the different groups of inverse analysis are reported. For each group of inverse analysis the scores of its successful simulations are summed.
The third and seventh columns present the scores assigned to the groups of inverse analysis in the first phase of the calibration process. The best score is the highest value while the worst score is the lowest value and a chromatic scale from green (best score) to red (lower score) helps in visualizing the results. The fourth and eighth columns report the excluded groups of inverse analysis whose indirect simulations do not overcome the 1st calibration phase. The 7 groups of inverse analysis excluded in the preliminary phase of the calibration process (see chapter §7_2) have been included in the analysis and are coloured in red (1st column) (1_1_36; 1_1_41; 1_1_45; 1_1_47; 1_1_42; 1_1_46 and 1_1_37).

As it is possible to observe the score assigned to these groups confirms again their poor performance also in the first calibration phase. The classification carried out in the first phase is more stringent than the one operated in the preliminary phase and globally the performance of the indirect simulations seems poorer in this phase. In fact, a part from the 7 groups of inverse simulations excluded in the preliminary phase, problems in the performance of other groups of simulations are individuated in the first phase of the calibration (as for VGM_fluvial_wc_w=TMA; VGM+2cm_foundation_PH_w=1 and VGM+2cm_aquifer_PH_w=1).
This is due to the fact that this phase looks at the performance of the simulations in the peak and off-peak periods separately, giving more ''weight'' to the performance during flood events while in the preliminary phase only the global behaviour of the simulation was under scrutiny. As it is possible to observe in Table 31:

- The fluvial layer could be optimized with good results using any investigated hydraulic model and any weighting distribution (w=1 and w=TMA).
- The foundation layer shows difficulties in the optimization of its hydraulic parameters in all the analysed casuistries (w=1 or w=TMA using HVGM, VGM or VGM-2cm).
- The indirect simulations that optimize the aquifer layer (saturated permeability) show a performance sufficiently good (HVGM shows the highest score).
- The indirect simulations optimizing the embankment layer show the best performance and with great probability, this is due to the fact that an higher number of sensors (monitoring wc and pwp) is installed in this layer with respect to the other layers. A greater number of observation points of high-quality has, with absolute certainty, a relevant impact on the inverse analysis performance, leading to a better set of optimized parameters.
- It is important to highlight that the groups of inverse analysis 1_1_22 and 1_1_25 that optimize the hydraulic parameters of three layers together show a very good performance and this is a very positive fact because a good calibration programme should aim to enhance the predictions of as many layers as possible in multi-layered domains.
- As could be observed, 23 groups of inverse analysis (over 33) overcome the first calibration phase (rejection of 10 groups).
- Looking at the groups of inverse analysis that do not overcome the first calibration phase, it is possible to observe that 8/10 groups use a TMA weighting distribution and the remaining 2/10 groups use the hydraulic model VGM-2cm. As observed also in Table 29, in general a poorer performance of the indirect simulations is observed for the weighting distribution TMA (with respect to all weights=1) and the hydraulic model VGM-2cm (with respect to VGM and HVGM).

The representation of Table 31 allows to individuate the groups of inverse analysis that improve the most the model performance. Due to the high number of investigated indirect simulations, it is often difficult or impossible to look at the results of each single simulation and be able to draw general conclusions on the calibration programme, f.e where the major problems have been detected in the simulation performance looking at the values of the set of metrics.

GROUP OF INVERSE ANALYSIS	IDENT. CODE	SCORE_1st CALIBRATION PHASE	ACCEPTANCE
HVGM_emb_PH_w=1	1_1_1	8,7	ACCEPTED
HVGM_emb_WC_w=1	1_1_2	11,1	ACCEPTED
HVGM_emb_PH_w=1	1_1_3	8,8	ACCEPTED
HVGM_emb_WC_w=1	1_1_4	10,5	ACCEPTED
HVGM_emb_PH_w=TMA	1_1_36	5,2	REJECTED
HVGM_fluv_WC_w=1	1_1_7	9,7	ACCEPTED
HVGM_fluv_WC_w=TMA	1_1_39	8,6	ACCEPTED
HVGM_found_PH_w=1	1_1_9	7,7	ACCEPTED
HVGM_found_PH_w=TMA	1_1_41	3,1	REJECTED
HVGM_aquifer_PH_w=1	1_1_17	7,7	ACCEPTED
HVGM_emb+found_PH_w=1	1_1_13	9,0	ACCEPTED
HVGM_emb+found_PH_w=TMA	1_1_45	4,7	REJECTED
HVGM_emb+found+fluv_PH+WC_w=1	1_1_15	11,1	ACCEPTED
HVGM_emb+found+fluv_PH+WC_w=TMA	1_1_47	4,5	REJECTED
HVGM_emb+fluv_wc_w=1	1_1_11	12,8	ACCEPTED
VGM_emb_PH_w=1	1_1_5	8,6	ACCEPTED
VGM_emb_WC_w=1	1_1_6	11,0	ACCEPTED
VGM_fluv_WC_w=1	1_1_8	9,6	ACCEPTED
VGM_found_PH_w=1	1_1_10	6,7	ACCEPTED
VGM_found_PH_w=TMA	1_1_42	3,10	REJECTED
VGM_aquifer_PH_w=1	1_1_18	8,2	ACCEPTED
VGM_emb+found_PH_w=1	1_1_14	9,0	ACCEPTED
VGM_emb+found_PH_w=TMA	1_1_46	4,6	REJECTED
VGM_emb+found+fluv_PH+WC_w=1	1_1_16	11,0	ACCEPTED
VGM_emb+fluv_WC_w=1	1_1_12	9,4	ACCEPTED
VGM_emb_PH_w=TMA	1_1_37	5,6	REJECTED
VGM_emb+fluv_WC_w=TMA	1_1_44	9,6	ACCEPTED
VGM_fluv_WC_w=TMA	1_1_40	5,6	REJECTED
VGM+2cm_embank_WC_w=1	1_1_20	11,2	ACCEPTED
VGM+2cm_fluvial_WC_w=TMA	1_1_56	9,5	ACCEPTED
VGM+2cm_foundation_PH_w=1	1_1_26	5,2	REJECTED
VGM+2cm_emb+fluv_PH_w=1	1_1_28	10,0	ACCEPTED
VGM+2cm_aquifer_PH_w=1	1_1_34	6,5	REJECTED

Table 31: Scores attributed to each group of inverse analysis in the first phase of the calibration. A coloured scale helps the modeller to visualize the different performance (green for the higher performance, red for the poorer performance). It is relevant to report that the groups of simulations 1_1_11; 1_1_12; 1_1_37; 1_1_40; 1_1_44 are composed of a number of successful simulations ≤3.

Table 32 helps in understanding where the considered groups of inverse analysis show problems in the fitting between observed and simulated data. To do so, the behaviour of the different ''group of metrics'' (accuracy metrics, efficiency metrics and model selection criteria) is investigated for each group of inverse analysis by using our scoring strategy.

The considered ''groups of metrics'', whose points are summed up, are the following:

- the accuracy metrics for the whole period: RMSE, MAE, R^2.
- -the efficiency metrics for the whole period : NSE, NSE_{rel}, NSE_j, AI'.
- the model performance metrics for the whole period: AIC, BIC, HQIC.
- the efficiency metrics for the flood peak periods :NSE, NSE_{rel}, NSE_j, AI'.
- the accuracy metrics for the flood peak periods: RMSE, MAE, R^2.
- the efficiency metrics for the off-peak periods: NSE, NSE_{rel}, NSE_j, AI'.
- the accuracy metrics for the off-peak periods :RMSE, MAE, R^2.

A chromatic scale has been used for the ease of the table reading: green for the maximum score of each group of metrics and red for the lowest score. As it is possible to observe the efficiency metrics show lower scores (poorer performance of the simulations) with respect to the accuracy metrics, in particular the efficiency metrics of the off-peak periods which are close to the lower limit of acceptance. Table 32 helps understanding where the problem related to a low performance of a group of metric stays f.e. the groups

of inverse analysis that optimize the foundation and aquifer layers show problems in the simulation of the flood peak periods (see in particular the values of the accuracy metrics for the foundation layer and the values of the efficiency metrics for the aquifer layer). As observed also in the preliminary phase (chapter §7_2), in general, the weighting distribution TMA shows a lower performance with respect to the simple and uniform weighting distribution equal to unity (w=1). The model performance evaluation metrics have scores closer to the lower edge for almost all the groups of inverse analysis that use the TMAT (only exception VGM-2cm_fluvial_wc_weight=TMA).

GROUPS OF INVERSE ANALYSIS	IDENT. CODE	RMSE, MAE, R^2	NSE, NSErel, NSEj AI'	AIC BIC HQIC	NSE, NSErel, NSEj AI'	RMSE, MAE, R^2	NSE, NSErel, NSEj AI'	RMSE, MAE, R^2
		ACCURACY METRICS-WHOLE PERIOD	EFFICIENCY METRICS-WHOLE PERIOD	MODEL PERFORMANCE EVALUATION- WHOLE PERIOD	EFFICIENCY METRICS-PEAK PERIODS	ACCURACY METRICS-PEAK PERIODS	EFFICIENCY METRICS OFF-PEAK PERIODS	ACCURACY METRICS OFF-PEAK PERIODS
HVGM_emb_PH_w=1	1_1_1	1,4	0,83	0,54	1,29	1,89	0,43	0,84
HVGM_emb_WC_w=1	1_1_2	1,8	0,30	1,00	0,66	4,54	0,07	1,43
HVGM_emb_PH_w=1	1_1_3	1,4	0,83	0,54	1,29	1,93	0,43	0,84
HVGM_emb_WC_w=1	1_1_4	1,6	0,27	1,00	0,59	4,23	0,06	1,43
HVGM_emb_WC_w=TMA	1_1_36	0,8	0,37	0,00	0,63	1,43	0,21	0,48
HVGM_fluv_WC_w=1	1_1_7	1,9	0,50	1,00	0,39	3,63	0,05	1,04
HVGM_fluv_WC_w=TMA	1_1_39	1,8	0,26	1,00	0,11	3,62	0,00	1,01
HVGM_found_PH_w=1	1_1_9	0,6	0,37	1,00	0,63	0,45	0,15	0,42
HVGM_found_PH_w=TMA	1_1_41	0,0	0,42	0,00	0,62	0,48	0,15	0,43
HVGM_aquifer_PH_w=1	1_1_17	1,4	0,37	1,00	0,00	2,01	0,05	1,00
HVGM_emb+found_PH_w=1	1_1_13	1,4	0,75	0,50	1,32	1,87	0,28	0,78
HVGM_emb+found_PH_w=TMA	1_1_45	0,9	0,31	0,25	0,70	1,19	0,15	0,48
HVGM_emb+found+fluv_PH+WC_w=1	1_1_15	1,6	0,92	0,50	1,36	2,68	0,45	1,14
HVGM_emb+found+fluv_PH+WC_w=TMA	1_1_47	1,0	0,29	0,00	0,68	1,27	0,14	0,48
HVGM_emb+fluv_wc_w=1	1_1_11	1,8	0,15	0,00	0,34	0,64	0,07	0,24
VGM_emb+fluv_wc_w=1	1_1_12	1,3	0,61	0,67	1,03	3,13	0,29	0,95
VGM_emb_PH_w=TMA	1_1_37	1,0	0,42	0,00	0,66	1,43	0,19	0,48
VGM_emb+fluv_wc_w=TMA	1_1_44	17,8	0,40	1,00	0,71	3,70	0,23	1,15
VGM_emb_PH_w=1	1_1_5	1,4	0,81	0,54	1,28	1,89	0,43	0,83
VGM_emb_WC_w=1	1_1_6	1,5	0,33	1,00	0,64	4,25	0,09	1,43
VGM_fluv_WC_w=1	1_1_8	2,0	0,52	1,00	0,33	3,70	0,00	1,01
VGM_found_PH_w=1	1_1_10	0,5	0,37	1,00	0,64	0,47	0,15	0,40
VGM_found_PH_w=TMA	1_1_42	0,0	0,42	0,00	0,62	0,48	0,15	0,43
VGM_aquifer_PH_w=1	1_1_18	0,9	0,37	1,00	0,08	1,81	0,03	0,78
VGM_emb+found_PH_w=1	1_1_14	1,4	0,75	0,50	1,32	1,87	0,29	0,78
VGM_emb+found_PH_w=TMA	1_1_46	0,9	0,33	0,00	0,74	1,14	0,15	0,48
VGM_emb+found+fluv_PH+WC_w=1	1_1_16	1,6	0,92	0,50	1,35	2,64	0,45	1,12
VGM_fluv_wc_w=TMA	1_1_40	1,0	0,42	0,00	0,66	1,43	0,19	0,48
VGM+2cm_embankment_WC_weight=1	1_1_20	1,8	0,29	1,00	0,50	4,28	0,04	1,43
VGM+2cm_fluvial_WC_weight=TMA	1_1_24	1,9	0,46	1,00	0,36	3,67	0,01	1,00
VGM+2cm_foundation_PH_weight=1	1_1_26	0,5	0,37	1,00	0,56	0,67	0,16	0,49
VGM+2cm_emb+fluv_PH_w=1	1_1_28	2,0	0,13	1,00	0,16	4,22	0,00	1,42
VGM+2cm_aquifer_PH_weight=1	1_1_34	0,9	0,37	1,00	0,17	1,94	0,03	0,79
MAX SCORE		2,00	1,01	1,00	1,43	4,28	0,48	1,44

Table 32: The table presents the score attributed to each group of metrics (accuracy, efficiency, model selection criteria) for each group of inverse analysis in order to visualize easily which groups show problems in the fitting between observed and simulated datasets.

In Table 33 the attention is focused on the single indirect simulations (only a small part of the 193 analysed indirect simulations are presented here for a sake of brevity). The scores of each ''group of metrics'' (accuracy, efficiency, model selection criteria) are reported here for each indirect simulation. Some groups of metrics do not show a change in the score between the simulations, this means that variations in the values of the metrics are small enough to not change the score associated to the metrics. The positive aspect is that at least one group of metrics always shows changes within the same group of inverse analysis. The groups of metrics that show changes within the same group of inverse analysis are coloured in blue. In the last column the final score obtained summing the score of each group of metrics has been reported.

Table 33 shows clearly why the mean of the values of a ''group of metrics'' is appropriate to describe the behaviour of a group of inverse analysis (as done in Table 31 and Table 32): the metrics of the simulations belonging to the same group are arranged in a narrow interval and the scores associated to these metrics accordingly differ only slightly. Nevertheless we have to remember that the three phases of the calibration

programme are applied to the single indirect simulation and the average behaviour of the groups of inverse analysis is investigated for a need of general overview on the performance of the whole set of indirect simulations and for a need of reporting the main result of each phase in a comprehensive manner.

The first phase of the calibration is based on a scoring system which has the purpose of eliminating from future phases the simulations that show a performance lower than a certain threshold.
The scoring method allows also a hierarchization of the indirect simulations (from the best to the worst) that is useful at the end of the calibration process when it is necessary to establish the final set of optimized parameters for the investigated hydraulic models. But, as seen in Table 33, simulations belonging to the same group of inverse analysis show often similar performances so in order to perform an univocal hierarchization, in phase 3 of the calibration, the scoring method will be substituted by a direct comparison of the performance of each simulation with the performance of the base simulation (which uses the initial set of hydraulic parameters). As already reported in the previous chapters, the initial set of hydraulic parameters is composed of the average value of each hydraulic property that come from the laboratory tests performed on the soils of the bank section (evaporation tests discussed in chapter §3_2). In this way, the enhancements in the model performance introduced by every indirect simulation will be individuated by comparison with a reference ''performance''. The set of hydraulic parameters that show the greater quantitative enhancement with respect to the ''reference simulation'' represents the final result of the whole calibration programme.

GROUPS OF INVERSE ANALYSIS		RMSE, MAE, R^2	NSE, NSErel, NSEj AI'	AIC, BIC, HQIC	NSE, NSErel, NSEj AI'	RMSE, MAE, R^2	NSE, NSErel, NSEj AI'	RMSE, MAE, R^2	
		ACCURACY METRICS-WHOLE PERIOD	EFFICIENCY METRICS-WHOLE PERIOD	MODEL PERFORMANCE EVALUATION-WHOLE PERIOD	EFFICIENCY METRICS-PEAK PERIODS	ACCURACY METRICS-PEAK PERIODS	EFFICIENCY METRICS OFF-PEAK PERIODS	ACCURACY METRICS OFF-PEAK PERIODS	TOT SCORE
PH__HVGM_ EMBANKMENT									
1_3_2	k	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
1_3_4	θs	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
1_3_5	n	1,375	0,833	0,500	1,310	1,697	0,441	0,813	6,969
1_3_7	αwet	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
1_3_9	k,α	1,375	0,833	0,500	1,310	1,913	0,441	0,813	7,185
1_3_10	k,θs	1,375	0,833	0,500	1,310	1,745	0,441	0,813	7,017
1_3_11	k, n	1,375	0,833	0,500	1,310	1,697	0,441	0,813	6,968
1_3_12	α,θs	1,375	0,833	0,500	1,310	1,697	0,441	0,813	6,969
1_3_13	α,θs,k	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
1_3_14	α,θs,k,n	1,375	0,833	0,500	1,310	1,913	0,441	0,813	7,185
1_3_15	θr	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
1_3_16	θr,k	1,375	0,833	0,500	1,310	1,697	0,441	0,813	6,969
1_3_17	α,θs,k,n, θr	1,375	0,833	0,500	1,310	1,913	0,441	0,813	7,185
1_3_13	θm, kwet, αwet	1,375	0,833	0,500	1,310	1,865	0,441	0,813	7,137
VWC_HVGM_EMBANKMENT									
1_4_1	k	1,500	0,125	0,999	0,476	4,392	0,032	1,425	8,949
1_4_2	α	2,000	0,375	0,999	0,774	4,690	0,106	1,425	10,369
1_4_3	θs	1,500	0,333	0,999	0,585	4,690	0,080	1,425	9,613
1_4_4	θr	1,500	0,125	0,999	0,448	4,690	0,026	1,425	9,213
1_4_5	n	1,500	0,250	0,999	0,448	4,690	0,026	1,425	9,338
1_4_6	k,α	1,500	0,417	0,999	0,816	4,392	0,090	1,425	9,638
1_4_7	k, θs	1,500	0,375	0,999	0,758	4,392	0,102	1,425	9,551
1_4_17	αw	1,500	0,125	0,999	0,448	4,690	0,026	1,425	9,213
PH_VGM_EMBANKMENT									
1_5_2	k	1,375	0,833	0,4995	1,302	1,697	0,441	0,799	6,947
1_5_3	α	1,375	0,833	0,4995	1,302	1,697	0,441	0,799	6,947
1_5_4	θs	1,375	0,833	0,4995	1,294	1,697	0,441	0,799	6,939
1_5_5	n	1,375	0,833	0,4995	1,303	1,865	0,441	0,799	7,115
1_5_9	k,α	1,375	0,833	0,4995	1,302	1,865	0,441	0,799	7,115
1_5_5	k,θs	1,375	0,833	0,4995	1,302	1,697	0,441	0,799	6,947
1_5_11	k, n	1,375	0,833	0,4995	1,302	1,865	0,441	0,799	7,115
1_5_12	α,θs	1,375	0,833	0,4995	1,302	1,745	0,441	0,799	6,995
1_5_13	α,θs,k	1,375	0,833	0,4995	1,303	1,745	0,441	0,799	6,996
1_5_14	α,θs,k,n	1,375	0,667	0,4995	1,303	1,697	0,441	0,799	6,781
1_5_16	θr,k	1,375	0,833	0,4995	1,302	1,689	0,441	0,799	6,939
1_5_17	α,θs,k,n, θr	1,375	0,833	0,4995	1,303	1,697	0,441	0,799	6,948
VWC_VGM_EMBANKMENT									
1_6_3	α	2,000	0,375	0,999	0,774	4,690	0,106	1,425	10,369
1_6_4	θs	1,500	0,333	0,999	0,585	4,690	0,080	1,425	9,613
1_6_5	n	1,500	0,250	0,999	0,448	4,690	0,026	1,425	9,338
1_6_9	k,α	1,500	0,292	0,999	0,815	4,690	0,107	1,425	9,828
1_6_10	k,θs	1,000	0,417	0,999	0,573	4,224	0,106	1,425	8,744
VWC_HVGM_FLUVIAL									
1_7_2	k	2,000	0,750	0,999	1,178	3,361	0,241	1,133	9,663
1_7_4	θs	2,000	0,458	0,999	0,199	3,739	0,000	1,033	8,428
1_7_10	k,θs	2,000	0,541	0,999	0,280	3,578	0,000	1,033	8,430
1_7_12	α,θs	2,000	0,625	0,999	0,313	3,746	0,000	1,015	8,698
1_7_15	θr	2,000	0,125	0,999	0,001	3,715	0,000	0,983	7,823
VWC_VGM_FLUVIAL									
1_8_4	θs	2,000	0,458	0,999	0,313	3,739	0,000	1,000	8,509
1_8_10	k,θs	2,000	0,458	0,999	0,313	3,739	0,000	0,983	8,492
1_8_12	α,θs	2,000	0,541	0,999	0,313	3,720	0,000	1,015	8,589
1_8_13	α,θs,k	2,000	0,625	0,999	0,389	3,596	0,000	1,023	8,631

Table 33: Examples of the scores attributed to each indirect simulation belonging to a certain ''group of inverse analysis'' in the first phase of the calibration. In blue the scores that show a change within the same ''group of inverse analysis''. In the last column on the right side, the sum of the scores for each indirect simulation.

7_3_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

In the first phase of the proposed methodological approach, the performance of the single indirect simulation is evaluated while, in the preliminary phase (chapter 7_2), the performance of groups of inverse analysis was under scrutiny. The first phase focuses on the differences between observed-simulated datasets by means of a selected set of metrics/indices. A scoring-base methodology is proposed to evaluate the performance of the indirect simulations. A set of metrics of typology accuracy, efficiency and model selection criteria has been chosen. The performance of each indirect simulation has been investigated separately in the flood peak periods, in the off-peak periods (drawdowns) and globally on the whole 2018 simulation period. Points are given to the metrics based on the closeness to the best fit. Points are scaled for each simulation according to these weights:

-each metric has an ''internal weight'' which is the weight given with respect to the other metrics of the same group (accuracy or efficiency or model selection criteria group)

-each metric has an ''external weight'' which is the weight of the group of metrics it belongs to with respect to the other groups (accuracy, efficiency and model selection criteria groups).

-flood peak and off-peak periods have different weights based on their persistence and hydrometric level reached.

The maximum score associated with the flood peak periods is 7,13; the one associated with the off-peak periods is 2,38 and the score associated with the whole simulation period is 5,5. The higher score attributed to the peak periods is due to the greater relevance of reliable pwp distributions during flood events to be used for possible future stability analysis. Using a scoring system to evaluate indirect simulations allows to obtain a ranking (from the best to the worst) of the available indirect simulations which allow to identify the best optimized set/s of parameters. The threshold score of acceptance of the indirect simulations is 7,5/15. 80% of the initial number of simulations has overcome the first phase of the calibration and will undergo the second phase. The first phase of the calibration process has been applied also to the groups of indirect simulations excluded in the preliminary phase in order to have a double check on the results obtained. These groups have obtained an insufficient score also according to the criteria of phase 1 which are more stringent with respect to the ones adopted in the pre-calibration phase. While the pre-calibration phase fixes only minimum thresholds, the first calibration phase allows a quantitative comparison between the investigated indirect simulations.

It has been observed that the HVGM with a mixed dataset and the VGM with a PH dataset assure the higher percentage of simulations that overcome the first calibration phase. The VGM-2cm assures with any typology of dataset and any weighting distribution the lower percentage of successful simulations which means a poorer performance. The performance of the simulations which use a wc dataset is in all cases poorer with respect to the same simulations which use a PH or a mixed dataset. The same could be observed for the indirect simulations that use the TMA weighting distribution with respect to a weighting distribution equal to unity.

Moreover it has been observed that HVGM and VGM have difficulties in the optimization of the hydraulic parameters of the foundation layer and in the optimization of groups of layers (embankment+foundation and embankment+foundation+fluvial) due probably to the lower number of observation points/sensors located in the foundation layer with respect to the other layers. For the same reason the optimization of

groups of layers that include the foundation shows a poor performance. The indirect simulations optimizing the embankment layer show the best performance and with great probability, this is due to the fact that an higher number of sensors (monitoring wc and pwp) is installed in this layer with respect to the other layers. A greater number of observation points of high-quality has, with absolute certainty, a relevant impact on inverse analysis performance, leading to the identification of a set of optimized parameters with a better performance. No successful simulations optimizing 2 or more layers are available for the HVGM-2cm, which assures in all cases a poorer performance compared to HVGM and VGM.

In chapter 7_4, the indirect simulations that have overcome the first calibration phase will be tested in the second phase of the calibration which focuses on the results of the inverse analysis (the optimized parameters) looking closely to the last iteration of the model.

7_4 PERFORMANCE EVALUATION OF INDIRECT SIMULATIONS: THE SECOND PHASE OF THE CALIBRATION PROCESS

In the second phase of the calibration process, the focus is moved to the optimized parameters and the goodness of them. The attention is restricted to a smaller group of simulations (thanks to the previous two phases) to allow an in-depth analysis of the indirect problem output. The indirect simulations that do not fulfil the threshold imposed in the first calibration phase are removed from the set of analysis to be investigated in the second phase.
Due to the fact that the goal of the second phase of the calibration process is different from the first one, other indices and metrics are used in the evaluation.

The first phase of the calibration process focuses on the quantification of the departure of the model output from observed and experimental measurements using various statistical and efficiency-based indices/indicators and statistic tests to judge the model performance. The second phase of the calibration process focuses on the results of the inverse analysis, on the optimized parameters, looking closely at the last iteration of the inverse analysis, which types of calculation error have been made, the user-defined tolerance to which the iteration process stops, the level of correlation between the optimized parameters, the range of values that is likely to include the true value of the parameter with a certain degree of confidence. A good fitting between simulated-observed datasets cannot always guarantee a good estimation of the parameters in the indirect problem, from this the necessity to perform also the second phase of the calibration process.

The adopted indices for the second phase are the following:

-**the 95% Confidence Interval**: The 95% confidence interval is calculated based on the evaluation of the objective function to assess its sensitivity to a particular parameter in its minimum. The confidence interval thus reflects the quality of the input dataset and the sensitivity of the model to the optimized parameter/s. The precision of the estimated parameters is assessed through confidence interval under certain hypothesis: the model error is zero, uncertainties are due only to measurements errors, inverse solution is converged to global minimum. Under these assumptions it is possible to estimate the 95% confidence limits and the parameter standard deviation value.
For each indirect simulation, the 95% confidence interval Δ between the lower and upper confidence limit of each optimized parameter is computed. Then the Range of Variation (ROV) of the parameter in percentage is calculated as follows [equation 139]:

$$\mathrm{ROV} = \frac{\Delta}{\text{parameter value}} * 100 \qquad \text{[equation 139]}$$

The range of variation is computed for each optimized parameter of the simulation then the score is given according to the amplitude of the range of variation. If Δ is narrow, then the fit is sensitive to this parameter and other values will not produce as good a fit.

-**the Mass Balance Error (ME)**: It gives an indication of the accuracy with which the finite difference matrix equations are being solved. Usually a value of mass balance below 1% is considered accurate. Score is given to each indirect simulation according to the value of mass balance error of the last iteration of the inverse problem. The lower the mass balance error, the greater the score given to the simulation. A necessary but not sufficient condition for solution accuracy stays in the ability of a numerical model to conserve the mass. A model that shows a poor mass balance behaviour is affected by uncertainties in the model predictions.

-**the Correlation Coefficient of the Parameters (CC)**: parameters correlation means that a group of parameters are mathematically related to each other in the model equations through some implicit functions. A high correlation between two parameters means that the parameters cannot be uniquely estimated, in other words more than a single parameter leads to the same response of the model (more than one combination of the parameters values produces a good fit). If the correlation between two parameter is low, it means that each parameter acts independently and is not influenced by the value of the other parameter.
A score is given to each couple of optimized parameters according to the value of its correlation coefficient. Then for each indirect simulation, the lower score (attributed to the couple with the higher correlation) is considered as representative of the whole simulation. This is penalizing for the indirect simulations that show few couples of correlated parameters but highlight the performance of simulations with no correlations or very low correlations between the couple of parameters.
According to Clay (2008) a value of correlation between 0,6-0,8 means a strong correlation, a value >0,8 means a very strong correlation. If model parameters are not well determined, model predictions are not well determined too, from all this the necessity to evaluate which parameters are identifiable thus inferring which model predictions are feasible. A high correlation manifests itself with a slowdown in the convergence rate, an increase in non-uniqueness and parameter uncertainty.
If a strong correlation between couple of parameters is detected, it means that the parameters could not be optimized together but it is necessary to determine them independently or to change the input dataset of observation points (Simunek et al, 2012).

There are two types of correlation parameters:

-**structurally non-identifiable parameters**: this correlation is related to the model structure independently of the experimental data. The set of structurally non identifiable parameters may be varied without changing the observation dataset, hence keeping the objective function constant. This means that these parameters are not uniquely identified. Confidence intervals of a structurally non identifiable parameter are infinite [-∞, +∞] in the logarithmic parameter space while structural identifiable parameters have always finite confidence limits. Due to the fact that structurally non-identifiable parameters are independent of the experimental dataset accuracy, the correlation could not be resolved reducing measurement errors or changing the input dataset. No work could be done by the modeller to resolve the non-identifiability of these parameters (Raue et al, 2009).

-**practical non-identifiable parameters**: the non-identifiability could be remediated by data improvement. This type of correlation arises when the amount and quality of the experimental data is insufficient.
Hydrus calculates ''asymptotic confidence intervals'' derived from the curvature of the likelihood function, the Hessian matrix. From this type of confidence intervals it is not possible to infer practical non-identifiability (Raue et al, 2009). As done in Hydrus inverse problem, practical non-identifiable parameters are detected using the Jacobian matrix that is determined for only the optimum parameters set with which

the correlation matrix is built describing the behaviour of the objective function in its minimum (Hopmans et al, 2002).
A graphical way to visualize the behaviour of the inverse problem is creating contour plots of the objective function φ against a couple of optimized parameters. Each point of the graph is obtained solving the flow equation for many combination of the couple of optimized parameter within a certain variation range, while the rest of the parameters are constant. Due to the fact that just couples of parameters are investigated, a 2D parameter space is obtained as cross section of the full parameter space. To study the behaviour of the whole parameter space it is necessary to calculate the response surface for all the possible couples of parameters. Only a partial evaluation is possible looking at few response surfaces. Response surfaces are able to show the presence of local minima, well defined global minimum, problems of parameter correlation and non-uniqueness of the solution (Hopmans et al, 2002).

In from Raue et, al (2009), three possibilities of response surfaces are presented. The white lines represents the 95% confidence regions, while the white star the optimal parameters estimation. Case A in which a structural non-identifiable couple of parameters results in a perfect flat valley infinitely extended along the corresponding functional relation between the couple of parameters. The confidence region is infinitely extended in both the directions. Case B represents practical non-identifiable couple of parameters showing a confidence region infinitely extended for θ_1 and $\theta_2 \rightarrow +\infty$. Case C presents the response surface of two identifiable parameters.

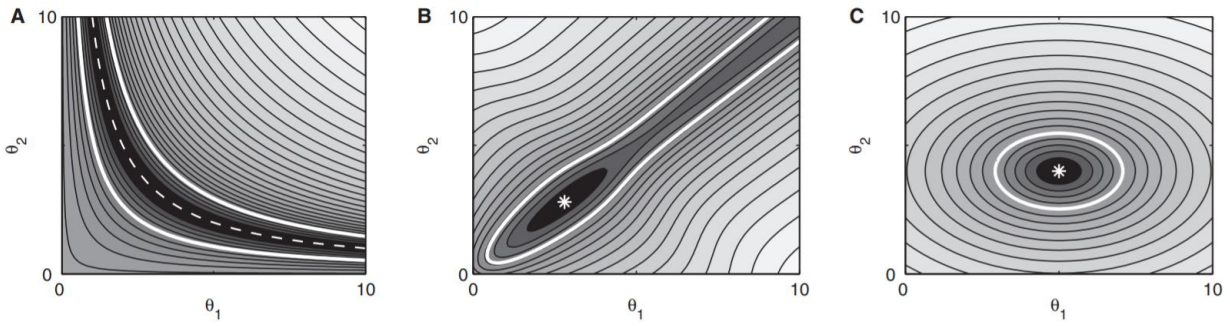


Figure 25: Three possibilities of response surfaces are presented: case A presents the contour plots of the objective function of a structural non-identifiable couple of optimized parameters, case B of a practical non-identifiable couple of parameters and case C of two identifiable parameters. In white the confidence intervals of each couple, the white star represents the optimal parameters estimation. Figure from Raue et al (2009).

-**the Simulation Tolerance (ST)**: the iteration process used to solve the Richards equation at each time step stops when the absolute change in pressure head (or water content) between two successive iterations is lower than the tolerance decided by the user. When convergence is not obtained in the indirect problem, the imposed tolerance is increased until reasonable values. Score is given to each simulation according to the value of the simulation tolerance used in the last iteration of the inverse analysis.

-**the Number of Optimized Parameters (NOP)**: a higher score is given to simulation that optimize more parameters at a time. The optimization of two or more parameters together is computationally more difficult for the code leading to problems of convergence or of parameters identifiability. Despite that, the optimization of more parameters at a time is extremely useful in order to investigate the combined effect of parameters on the simulation output. Usually problems of correlation between parameters increase increasing the number of optimized parameters.

-**ΔSSQ**: the difference between the sum of the square residual at the zero iteration (with the initial set of parameters) and at the last iteration (with the optimized set of parameters) is used in the second phase of the calibration to assess the enhancement of the model output in terms of observed-fitted dataset. The larger the ΔSSQ, the greater the improvement in the simulation output. SSQ is at the base of the objective

function, it is used to assess the agreement between experimental-observable datasets and to govern the inverse problem through its minimization. SSQ and ΔSSQ are calculated as follows [equation 140 and 141]:

$$SSQ = \sum_{i=1}^{n}(x_{i\,obs} - x_{i\,sim})^2 \quad \text{[equation 140]}$$

$$\Delta SSQ = (\sum_{i=1}^{n}(x_{i\,obs} - x_{i\,sim})^2)_{init.\ iter.} - (\sum_{i=1}^{n}(x_{i\,obs} - x_{i\,sim})^2)_{final\ iter.} \quad \text{[equation 141]}$$

where i=1..n, n is the number of observed data. Greater the ΔSSQ, greater the score attributed to each indirect simulation.

-**the Number of Iterations (NIT)**: a higher score is given to simulations that performed a higher number of iterations in the inverse analysis. In each iteration, Hydrus solves the governing equation using updated parameters, the values of the parameters become increasingly different from the initial dataset with each and every new iteration.

-**Standard Error Coefficient** (**SEC**): The Standard Error Coefficient gives an indication of how precisely the model estimates the unknown value of the optimized parameter. The smaller the SEC, the more precise the estimation. SEC is directly proportional to the standard error of the regression and inversely proportional to the square root of the sample size. This means that the use of a larger amount of data could possibly reduce the standard error. The standard error of the regression is calculated as follows [equation 142]:

$$s = \sqrt{\frac{1}{n-2}\sum_{i=1}^{n} e_i^2} \quad \text{[equation 142]}$$

where i=1…n , n is the sample size and e_i the error.
To assess if the estimation of the parameter is enough precise, it is necessary to compare SEC with the value of the parameter: if they have the same order of magnitude, it could be stated that the precision is poor. A more mathematical way to verify the degree of precision of the optimized parameter is using the statistic t-test or the p-value (see §2_7_2_4). Dividing the parameter for its standard error, the t-value is obtained. If the t-value is lower than the critical value obtained from the t-distribution table, it is not possible to declare a statistical difference (Ho is not rejected) that means that the parameter is not accurate and it could not be estimated from the equations system. The same conclusion could be obtained from the p-value when compared to the common levels of α (level of significance). In the second phase of the calibration, the p-value is calculated starting from the standard error coefficient for each optimized parameter of the simulation, then the max of the p-values of each simulation is computed. Score is given to the simulation according to this value. If the max p-value is lower than α (level of significance), the test is statistically significant (Ho rejected) and for this reason the maximum score is given. Moving away from α toward higher values, the score decreases.
In a summary of the scoring strategy for each metric considered in the second phase of the calibration is presented. The maximum score attributed to the best fit of each metric is 1 so the maximum score that an indirect simulation could achieve is 8. The lower acceptability score is 4.

Number of optimized parameters (NOP)		**Simulation tolerance (ST)**	
NOP=1	0	**ST≥0,5**	0
2≤NOP<4	0,5	**0,1≤ST≤0,5**	0,25
NOP≥4	1	**0,075≤ST≤0,1**	0,50
		0,05≤ST≤0,075	1
Mass error (ME)		**Number of iterations (NIT)**	
ME>0,1	0	**NIT=1**	0
0,05≤ME≤0,1	0,25	**1<NIT≤4**	0,25
0,01≤ME<0,05	0,50	**4<NIT≤8**	0,50
ME<0,01	1	**NIT>8**	1
ΔSSQ		**ROV (Range of Variation)**	
0≤ΔSSQ≤5	0	**0%≤ROV≤2,5%**	1

5<ΔSSQ≤10	0,25	**2,5%<ROV≤5%**	0,50
10<ΔSSQ≤15	0,50	**5%<ROV≤10%**	0,25
15<ΔSSQ≤20	0,75	**ROV>10%**	0
ΔSSQ>20	1		
Correlation coefficient (CC)		**p-value**	
0≤CC≤0,35 & -0,35≤CC≤0	1,0	**pv≤0,05**	1
0,35≤CC≤0,60 & -0,6≤CC≤-0,35	0,5	**0,05≤pv≤0,1**	0,50
0,60≤CC≤0,80 & -0,8≤CC≤-0,60	0,25	**0,1<pv≤0,2**	0,25
CC>0,8 & CC≤-0,8	0	**pv>0,2**	0

Table 34: Allocation of the scores for each metric/index considered in the second phase of the calibration process proportionally to the distance from the best fit of the metric/index.

For the presented case study, 53% (81) of the initial number of simulations (155) has been accepted :the sum of the scores of each indirect simulation is greater than 4 which is considered as the threshold. In , a summary of the number of the simulations that overcame the second phase of the calibration. As it is possible to observe, the HVGM is the hydraulic model with the highest number of successful simulations, while VGM+2cm has the lowest value, due to an overall difficulty in the convergence and a poor quality of results of the inverse analysis. The embankment layer has the highest number of successful indirect simulations due to a greater number of simulations performed with respect to the other layers and a larger dataset of observation points used in the inverse problem. In round brackets the percentage of simulations that have passed the second phase of the calibration with respect to the initial number of simulations falling into each category. The embankment layer and the groups of layers (embankment+fluvial; embankment+foundation; embankment+foundation+fluvial) show the higher percentages of indirect simulations that have passed the second calibration phase (percentages from 60 to 100%), the only exception is for the HVGM the groups of layers embankment+foundation+fluvial (20% of successful simulations). The VGM-2cm will be excluded from here on out because the restricted number of simulations that have passed the first phase, have shown an insufficient quality in the second phase. It is relevant to notice that the second phase of the calibration has proven to be more selective and stringent with respect to the first phase despite it was focused on an in-depth analysis of the observed-simulated dataset. In fact in the first phase, 155 indirect simulations over 193 passed the selection (80%), while in the second phase 81 over 155 (53%). Moreover it has to be remembered that the second phase of the calibration acts on a set of simulations already purged from the ones with a poor performance so the average performance is higher with respect to the first phase.

	HVGM	VGM	VGM-2cm
n° of indirect simulations for the optimization of the embankment layer	26 (60%)	12 (71%)	0
n° of indirect simulations for the optimization of the fluvial layer	2 (25%)	0	1 (33%)
n° of indirect simulations for the optimization of the foundation layer	4 (44%)	4 (33%)	0
n° of indirect simulations for the optimization of the aquifer layer	2 (22%)	6 (46%)	0
n° of indirect simulations for the optimization of the embankment+fluvial layers	2 (100%)	2 (100%)	0
n° of indirect simulations for the optimization of the embankment+foundation layers	6 (75%)	5 (71%)	0
n° of indirect simulations for the optimization of the	1 (20%)	8 (100%)	0

embankment+foundation+fluvial layers			
sum	39	36	1
total n° of successful simulations that have overcome the 2nd phase	**81**		
%successful simulations with respect to the n° of simulations that have overcome the 2nd phase	**53%**		

Table 35: Number of simulations that have passed the second phase of the calibration subdivided for layer and hydraulic model (HVGM, VGM, VGM-2cm). In round brackets the percentage of simulations that have passed the second phase with respect to the initial number of simulations falling into that category (f.e. the % of indirect simulations optimizing the embankment layer using a VGM is calculated with respect to the n° of simulations optimizing the embankment layer using the VGM that have overcome the first calibration phase).

In other statistics of the groups of inverse simulations that have overcome the second phase of the calibration are presented. Some conclusions could be drawn from the presented table:

- The groups of indirect simulations that use the TMAT have shown a very poor performance in both the calibration phases. The number of simulations which use the TMAT and which have overcome the second calibration phase is very reduced. Just 3 groups of inverse analysis have overcome the 2nd calibration phase (1_1_39 for the HVGM, 1_1_44 for the VGM and 1_1_24 for the VGM-2cm) and it is relevant to report that all these groups optimize the hydraulic parameters of the fluvial layer. We can conclude that the TMAT has not been able to guarantee good results in the indirect problem.
- Few groups of inverse analysis that optimize more than one single layer have overcome both the calibration phases: embankment+foundation_PH (HVGM, VGM) and embankment+foundation+fluvial_PH+wc (HVGM, VGM).
- The hydraulic model that shows the higher percentage of accepted simulations is the VGM: 63% with respect to the number of simulations of the same typology that have overcome the previous phase.
- The indirect simulations with a pressure head input dataset show the higher degree of acceptance (75% for the HVGM and 81% for the VGM).
- The indirect simulations with a water content input dataset show a low degree of acceptance (43% for the HVGM and 64% for the VGM). We can state that the pressure head dataset allows to obtain better results in the inverse estimation of hydraulic parameters in the presented case study.
- For the HVGM the optimization of the hysteretic parameters (θ_m, α_w, K_w) is extremely difficult and any simulation optimizing the parameters of the wetting curve have overcome the second calibration phase. This could be explained remembering the results obtained in chapter §6_1 in which the hysteretic behaviour of the bank section has been investigated and just limited effects due to the adoption of an hysteretic model to represent the river bank behaviour have been identified.
- Aquifer (HVGM), fluvial (VGM), foundation (VGM) and embankment+fluvial+foundation (HVGM) are the layers that have obtained the lower percentage of simulations that have overcome the second calibration phase. With great probability the availability of a lower number of observation points for the layers different from the embankment assures a poorer performance of the inverse analysis as could be read from the differences between the percentage of successful simulations of the embankment and of other layers (fluvial, foundation, aquifer). The combined optimization of 3 layers together results in an high complexity for the Levenberg- Marquardt algorithm that could explain the difficulties in the indirect simulations convergence for couple and triplet of layers. Maintaining a low number of fitting parameters, as in the majority of the groups of inverse analysis that have overcome both the calibration phases, leads to a better performance of the inverse analysis, in agreement with the findings of Simunek et al (2001) and Hopman et al (2002). In fact

the uncertainties are reduced when a limited number of hydraulic parameters are optimized. Moreover a decrease in the number of parameters assures to reduce the non-uniqueness of the solution in the inverse problem (Hopman et al, 2002).

- The hydraulic parameters of the fluvial layer have proven to be the most difficult to optimize. The HVGM allows the optimization of Ks, θs parameters of the fluvial layer while the VGM has no successful simulations. The combining optimization of the fluvial layer with other layers allows to obtain better results as for the group 1_1_16 (embankment+foundation+fluvial).
- Two groups of inverse analysis (1_1_8 and 1_1_20) are rejected in the second calibration phase due to the fact that none of the indirect simulations that belong to the groups show a sufficient performance. At the end of the second phase of the calibration, 20 groups of inverse analysis will be tested in the third calibration phase (11 groups HVGM, 8 groups VGM, 1 VGM-2cm, zero HVGM-2cm).

GROUPS OF INVERSE ANALYSIS	IDENT. CODE	N° OF SUCCESSFUL SIMULATIONS	% SUCCESSFUL SIMULATIONS	OPTIMIZED PARAMETERS
HVGM_emb_PH_w=1	1_1_1	11	0,85	n, k, α, θs, θr
HVGM_emb_WC_w=1	1_1_2	3	0,38	k, α, θr
HVGM_emb_PH_w=1	1_1_3	10	0,71	n, k, α, θs, θr
HVGM_emb_WC_w=1	1_1_4	2	0,25	k, α, θs
HVGM_fluv_WC_w=1	1_1_7	1	0,20	k, θs
HVGM_fluv_WC_w=TMA	1_1_39	1	0,33	θs
HVGM_found_PH_w=1	1_1_9	4	0,44	k, α, θs, θr
HVGM_aquifer_PH_w=1	1_1_17	2	0,22	k, α, θs, θr, n
HVGM_emb+found_PH_w=1	1_1_13	6	0,75	k, α, θs, θr
HVGM_emb+found+fluv_PH+WC_w=1	1_1_15	1	0,20	k, θr
HVGM_emb+fluv_w=1	1_1_11	2	1,00	α, θs
VGM_emb+fluv_wc_w=TMA	1_1_44	1	1,00	n, k, α, θs
VGM_emb_PH_w=1	1_1_5	10	0,83	n, k, α, θs, θr
VGM_emb_WC_w=1	1_1_6	2	0,40	k, α
VGM_emb+fluv_wc_w=1	1_1_12	1	1,00	θs
VGM_found_PH_w=1	1_1_10	4	0,33	n, k, α, θs, θr
VGM_aquifer_PH_w=1	1_1_18	6	0,46	n, k, α, θs, θr
VGM_emb+found_PH_w=1	1_1_14	5	0,71	n, k, α, θs, θr
VGM_emb+found+fluv_PH+wc_w=1	1_1_16	8	1,00	n, k, α, θs, θr
VGM_fluv_w.c._w=1	1_1_8	0	0,00	-
VGM+2cm_emb_wc_w=1	1_1_20	0	0,00	-
VGM+2cm_fluvial_WC_weight=TMA	1_1_24	1	0,33	n, k, α, θs

% of accepted simulations_VGM	0,63	% of simul_VGM_w=1	0,59
% of accepted simulations_HVGM	0,51	% ofsimul_VGM_w=TMA	1,00
% of accepted simulations_VGM+2cm	0,33	% of simul_HVGM_w=1	0,52
% of accepted simulations_HVGM_PH	0,75	% of simul_HVGM_w=TMA	0,33
% of accepted simulations_HVGM_wc	0,43	% of sim_VGM+2cm_w=1	0,00
% of accepted simul_HVGM_wc+PH	0,20	% of simul_VGM+2cm_w=TMA	0,33
% of accep. simulations_VGM_PH	0,81		
% of accep.simulations_VGM_wc	0,64		
% of accep simul._VGM_wc+PH	1,00		

Table 36: The table presents the n° of indirect simulations in each group of inverse analysis that have overcome the second calibration phase (third column) and the % of these simulations with respect to the number of simulations in the same group that have overcome the first calibration phase (fourth column). In the last column the hydraulic parameters that are optimized in each group (we have to remember that for the groups of inverse analysis that optimize more than one layer at a time, a set of optimized parameters is obtained for each layer). The groups of inverse simulations that present zero successful simulations (performance below the imposed threshold) will be excluded from the third calibration phase.

As done in the first phase of the calibration, the average score of the considered metrics are computed for each group of inverse analysis and reported in . As already mentioned, it is legitimate to consider the average points for each metric because the indirect simulations belonging to a group use the same observation dataset and optimize the hydraulic parameters of the same layer (but in different combination). Moreover the values of the metrics of the simulations belonging to a group of inverse analysis fall in a narrow range.
As it is possible to observe in :

- the groups of simulations with the higher score are the ones that optimize the embankment (HVGM), the embankment+foundation (HVGM and VGM), the embankment+fluvial (HVGM, VGM) the embankment+foundation+fluvial layers (VGM).
- the average scores of the groups of inverse analysis (ranging between 4-5 in the pre 2nd phase) are quite low compared to the maximum score (8). reports the score attributed to each group of inverse analysis after the elimination of the indirect simulations (post 2nd phase) that do not fulfil the imposed threshold, as could be observed all the groups show an enhancement of the performance as expected. The new range of the score attributed to the 2nd phase is between 4 – 6,3.
- as already mentioned, only four groups of inverse analysis have been excluded (zero successful simulations after the 2nd phase of calibration): 1_1_8 and 1_1_40 for the VGM and 1_1_20 and 1_1_28 for the VGM-2cm.
- a general low score has been observed for the mass error of the inverse analysis using the VGM, sign of a low solution accuracy (uncertainties in the prediction).
- a general low score attributed to NIT (number of iterations) for almost all the groups of inverse simulations suggests that the optimized sets of hydraulic parameters differ only slightly from the initial set of parameters due to a reduced number of performed iterations. In fact in each iteration of the inverse analysis, the value of the parameters to optimize are changed increasingly from their initial value.
- the high scores of the groups of inverse analysis attributed to the p-value suggest the ability of the system of equations to estimate singularly the optimized parameters (Ho rejected).
- the high scores attributed to the simulation tolerance of the groups of inverse analysis HVGM and VGM are proof of the precision with which the inverse problem has been solved. Similarly the low values of ST for the VGM+2cm simulations suggest a low accuracy of the obtained sets of optimized parameters.

GROUP OF INVERSE ANALYSIS	IDENT. CODE	NOP	ME	NIT	ST	ROV	p-value	CC	ΔSSQ	AVERAGE SCORE PRE 2nd PHASE	AVERAGE SCORE POST 2nd PHASE	ACCEPTANCE
HVGM_emb_PH_w=1	1_1_1	0,4	0,2	0,1	0,9	1,0	1,0	0,3	0,6	4,5	4,7	ACCEPTED
HVGM_emb_WC_w=1	1_1_2	0,6	0,4	0,3	0,5	0,4	1,0	0,1	0,5	3,8	4,0	ACCEPTED
HVGM_embankment_PH_w=1	1_1_3	0,4	0,1	0,1	0,9	0,9	0,9	0,4	0,5	4,3	4,8	ACCEPTED
HVGM_embankment_wc_w=1	1_1_4	0,1	0,3	0,3	0,8	0,6	1,0	0,2	0,0	3,3	4,0	ACCEPTED
HVGM_fluvial_wc_w=1	1_1_7	0,2	0,2	0,3	0,7	0,3	0,8	0,3	0,2	2,8	5,3	ACCEPTED
HVGM_fluvial_wc_w=TMA	1_1_39	0,0	0,1	0,3	0,6	0,5	1,0	0,0	1,0	3,5	4,3	ACCEPTED
HVGM_foundation_PH_w=1	1_1_9	0,3	0,2	0,1	0,8	0,6	0,9	0,4	0,6	3,9	4,8	ACCEPTED
HVGM_aquifer_PH_w=1	1_1_17	0,5	0,3	0,0	1,0	1,0	1,0	0,0	0,0	3,8	4,3	ACCEPTED
HVGM_emb+found_PH_w=1	1_1_13	0,8	0,1	0,2	0,9	0,5	1,0	0,6	0,5	4,6	4,7	ACCEPTED
HVGM_emb+found+fluv_PH+wc_w=1	1_1_15	0,6	0,2	0,1	0,9	0,6	0,8	0,6	0,2	3,9	6,3	ACCEPTED
HVGM_emb+fluv_w=1	1_1_11	0,5	0,3	0,3	1,0	0,5	1,0	1,0	0,3	4,8	4,8	ACCEPTED
VGM_emb+fluv_wc_w=TMA	1_1_44	0,5	0,0	0,0	0,0	1,0	1,0	1,0	1,0	4,5	4,5	ACCEPTED
VGM_embankment_PH_w=1	1_1_5	0,4	0,0	0,0	0,9	1,0	0,9	0,3	0,8	4,3	4,6	ACCEPTED
VGM_embankment_WC_w=1	1_1_6	0,2	0,1	0,4	0,5	0,7	1,0	0,3	1,0	4,1	4,3	ACCEPTED
VGM_emb+fluv_wc_w=1	1_1_12	0,5	0,0	0,3	0,3	1,0	1,0	1,0	0,3	4,3	4,3	ACCEPTED
VGM_foundation_PH_w=1	1_1_10	0,4	0,0	0,0	0,5	0,7	1,0	0,4	0,5	3,6	4,3	ACCEPTED
VGM_aquifer_PH_w=1	1_1_18	0,4	0,0	0,0	1,0	0,9	0,8	0,5	0,3	3,8	4,8	ACCEPTED
VGM_emb+found_PH_w=1	1_1_14	0,5	0,0	0,3	0,9	1,0	1,0	0,4	0,6	4,7	5,3	ACCEPTED
VGM_emb+found+fluvial_PH+wc_w=1	1_1_16	0,8	0,0	0,1	0,4	1,0	1,0	0,8	0,6	4,6	4,6	ACCEPTED
VGM_fluv_w.c._w=1	1_1_8	0,4	0,0	0,4	0,4	0,3	1,0	0,4	0,3	3,1	-	REJECTED
VGM_fluv_wc_w=TMA	1_1_40	0,0	0,0	0,5	0,3	0,0	0,0	0,0	0,0	0,8	-	REJECTED
VGM+2cm_emb+fluv_PH_weight=1	1_1_28	1,0	0,0	0,3	0,0	0,0	1,0	0,5	0,0	2,8	-	REJECTED
VGM+2cm_emb_wc_w=1	1_1_20	0,0	0,0	0,1	0,3	0,8	1,0	0,0	0,0	2,1	-	REJECTED
VGM+2cm_fluvial_wc_weight=TMA	1_1_56	0,5	0,0	0,4	0,3	0,6	1,0	0,6	1,0	4,4	4,6	ACCEPTED
MAX SCORE		1	1	1	1	1	1	1	1	8	8	

Table 37: The average score of each metric of the second phase of the calibration is computed for the different groups of inverse analysis. NOP stands for ''number of optimized parameters'', ME for ''mass error'', NIT for ''number of iterations'', ST for ''simulation tolerance'', ROV for ''Range of Variation'', CC for ''Correlation Coefficient'', ΔSSQ is the difference between the sum of the square residual at the zero iteration (with the initial set of parameters) and the last iteration (with the optimized set of parameters). In the eleventh column the average total score attributed to each group of inverse analysis (pre 2nd phase of calibration) and in the twelfth column the average total score post second phase (after the elimination of the indirect simulations that do not fulfil the minimum threshold).

In , comparing the score of the first calibration phase of each group of inverse analysis with the score of the second calibration phase, it is possible to observe a different evaluation of the model performance. A chromatic scale from green (best score) to red (lower score) is used to ease the table reading. The fifth column in reports the sum of the score of the first plus the second calibration phases.

- The greatest differences between the scores of the two calibration phases could be observed for the groups of inverse analysis HVGM_foundation_PH_w=1 (1_1_9); VGM_foundation_PH_w=1 (1_1_10) and VGM+2cm_emb_wc_w=1 (1_1_20). This is due to the fact that a good fit between the observed and simulated datasets in the first phase do not exclude problems of correlation between the optimized parameters, problems of non-identifiability of the parameters or low accuracy in the estimation (parameters and characteristics on which the second calibration phase focuses).
- The foundation layer using a VGM and HVGM and a weighting distribution equal to 1 presents the poorest results (12,5 for HVGM and 11,0 for VGM). A low performance could be detected also for the aquifer layer (12,0 for HVGM and 13,0 for VGM).
- Three groups of inverse analysis have shown an excellent performance in both the calibration phases (1^{st}+2^{nd}): HVGM_embankment+foundation+fluvial_wc+PH_w=1 (1_1_15); HVGM_fluvial_wc_w=1 (1_1_7) and HVGM_embankment+fluvial_wc_w=1 (1_1_11). This means that the hydraulic parameters of the embankment and fluvial layers could be estimated singularly or together with high accuracy by means of inverse analysis.

GROUPS OF INVERSE ANALYSIS	IDENT. CODE	AVERAGE SCORE 1st PHASE	AVERAGE SCORE 2nd PHASE	AVERAGE TOTAL SCORE	ACCEPTANCE
HVGM_emb_PH_w=1	1_1_1	8,7	4,7	13,4	ACCEPTED
HVGM_emb_WC_w=1	1_1_2	11,1	4,0	15,1	ACCEPTED
HVGM_embankment_PH_w=1	1_1_3	8,8	4,8	13,6	ACCEPTED
HVGM_embankment_WC_w=1	1_1_4	10,5	4,0	14,5	ACCEPTED
HVGM_fluvial_WC_w=1	1_1_7	9,7	5,3	15,0	ACCEPTED
HVGM_fluvial_WC_w=TMA	1_1_39	8,6	4,3	12,9	ACCEPTED
HVGM_foundation_PH_w=1	1_1_9	7,7	4,8	12,5	ACCEPTED
HVGM_aquifer_PH_w=1	1_1_17	7,7	4,3	12,0	ACCEPTED
HVGM_emb+found_PH_w=1	1_1_13	9,0	4,7	13,7	ACCEPTED
HVGM_emb+found+fluv_PH+WC_w=1	1_1_15	11,1	6,3	17,4	ACCEPTED
HVGM_emb+fluv_wc_w=1	1_1_11	12,8	4,8	17,6	ACCEPTED
VGM_emb+fluv_WC_w=TMA	1_1_44	9,6	4,5	14,1	ACCEPTED
VGM_embankment_PH_w=1	1_1_5	8,6	4,6	13,2	ACCEPTED
VGM_embankment_WC_w=1	1_1_6	11,0	4,3	15,3	ACCEPTED
VGM_emb+fluv_WC_w=TMA	1_1_12	9,4	4,3	13,7	ACCEPTED
VGM_foundation_PH_w=1	1_1_10	6,7	4,3	11,0	ACCEPTED
VGM_aquifer_PH_w=1	1_1_18	8,2	4,8	13,0	ACCEPTED
VGM_emb+found_PH_w=1	1_1_14	9,0	5,3	14,3	ACCEPTED
VGM_emb+found+fluvial_PH+WC_w=1	1_1_16	11,0	4,6	15,6	ACCEPTED
VGM_fluv_WC_w=1	1_1_8	9,6	3,1	12,7	REJECTED
VGM_fluv_WC_w=TMA	1_1_40	5,6	0,8	6,4	REJECTED
VGM+2cm_emb+fluv_PH_w=1	1_1_28	10,0	2,8	12,8	REJECTED
VGM+2cm_emb_WC_w=1	1_1_20	11,2	2,1	13,3	REJECTED
VGM+2cm_fluvial_WC_w=TMA	1_1_56	9,5	4,6	14,1	ACCEPTED
MAX SCORE		15	8	23	

Table 38: The scores of the first calibration phase are compared to the scores of the second calibration phase of each group of inverse analysis. The third column reports the sum of the scores of the 1st+2nd phases. In red the four groups of inverse analysis that do not overcome the second calibration phase.

As introduced before, the Response Surface Analysis is an effective way to investigate the behaviour of the objective function in proximity of the optimized couple of parameters and the well-posedness of the inverse problem. The Response Surface Analysis is able to detect the presence of local minima in proximity of a global minimum and any potential problems of parameter correlation and sensitivity. In , and , three examples of Response Surface Analysis applied to three couples of parameters (1st couple belonging to the embankment layer and the other two couples to the foundation layer) are presented.

presents the behaviour of the objective function in the 2D space of the optimized parameters Ks-θs belonging to the embankment layer (VGM). The couple of optimized parameters belongs to the group of inverse analysis 1_1_5 that showed a good performance both in the first and second calibration phases.

The space of the parameters investigated is between 0,396-0,399 for θs (using a step of 0,0005(-)) and $8{,}94*10^{-5}$ -$9{,}0*10^{-5}$ m/min for the saturated permeability K_s (using a step of $0{,}02*10^{-5}$ m/min).
The surface created by the values of the objective function (z-axis) in correspondence of the couple (Ks-θs) shows a clear global minimum. To obtain a smoother surface (no edges) it is only necessary to reduce the grid of investigation of the space of the parameters but it is computationally more demanding and for the purpose of this analysis not necessary. The optimized parameter θs is equal to 0,398 while the optimized saturated permeability of the embankment layer is equal to $8{,}96*10^{-5}$ m/min. The same information could be obtained from the 3D graph on the left and its projection on the vertical plain (x-z axis and y-z axis). No other local minima are detected in the investigated portion of the parameters space.
and represent two cases in which a defined global minimum has not been found in the optimization of the couple Ks-alfa and Ks-θs of the foundation layer (HVGM). As it is possible to observe from the 3D surfaces in , the objective functions decreases towards very low values of alfa (<0,02) which are unacceptable values for the investigated layer and far below the reliable values obtained by laboratory tests. The poor performance of the indirect simulations Ks-θs and Ks-α could be extended to all the simulations that belong

to the group of inverse analysis 1_1_9 (foundation_HVGM_PH_w=1). In fact, as could be observed in , the performance of this group is quite poor compared to the other groups both in the first and second calibration phases.

The inverse analysis in the two cases presented in and is not able to identify the correct optimized couple of parameters in the investigated portion of parameter space. Changing the input set of initial parameters used in the indirect problem or adding new reliable information to the dataset of observation points could be possible solutions to overcome this problem. The investigation of the objective function in the parameters space could be relevant when problems of correlation between parameters, non-uniqueness of the solution or the possibility of the existence of multiple minima for the same couple of parameters are detected. Due to the higher computational demand to perform a Response Surface Analysis instead of applying a set of metrics, only few couples of optimized parameters could be tested when problems are detected from the application of the set of metrics/indices in the 1^{st} and 2^{nd} phases of the calibration. In fact the Response Surface Analysis is able to corroborate (or not) hypothesis made by the modeller from the in-depth observation of the values assumed by the metrics.

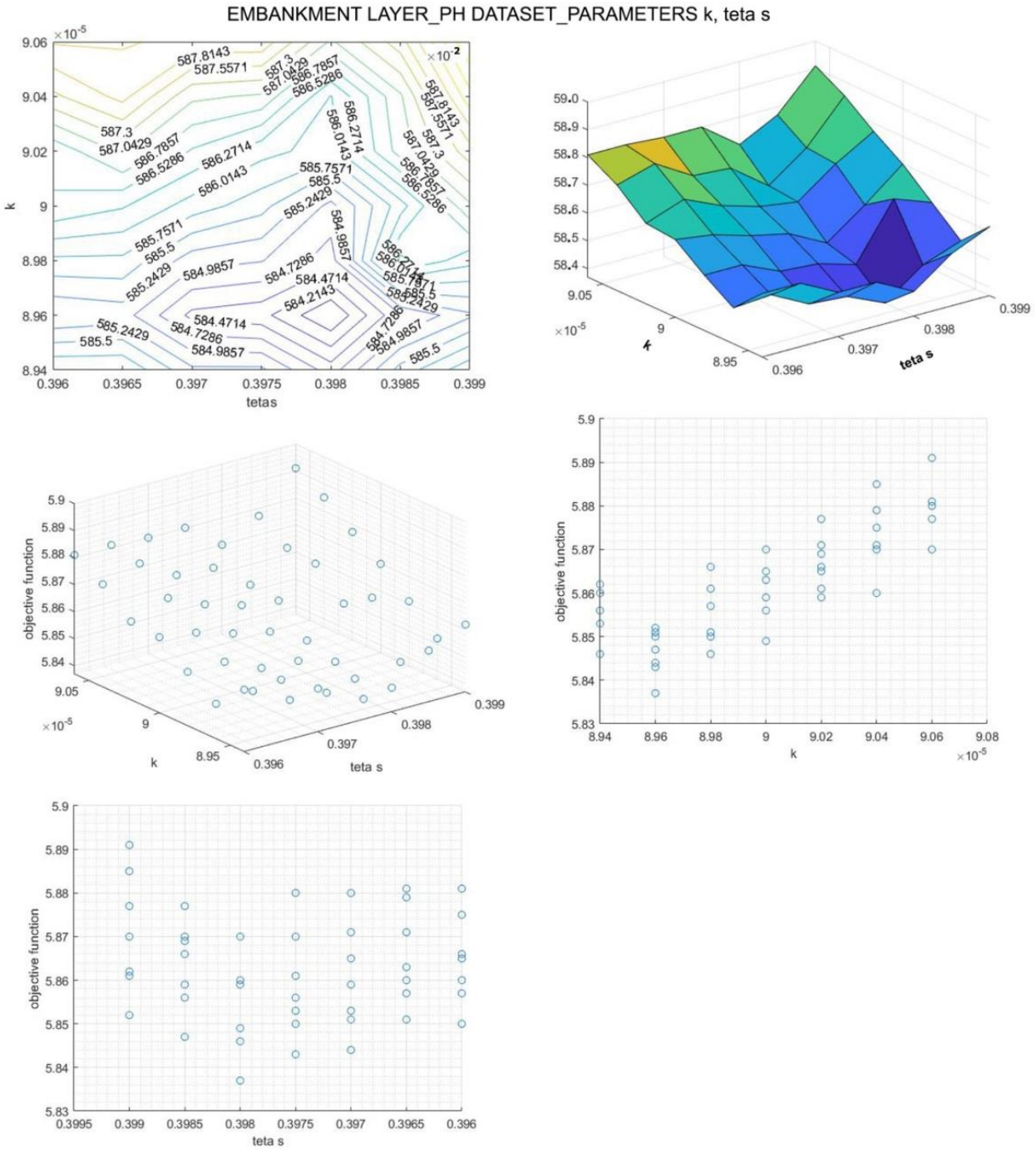


Figure 26: The behaviour of the objective function in the 2D/3D space of the optimized parameters k-θs belonging to the embankment layer (VGM) is presented. A clear local minimum is detected in the 3D surface that lies in the space k-θs-objective function (right side on the top) and in the 2D plain which shows the profiles of the objective function in the space k-θs (left side on the top). The figure in centre on the left represents the values of the objective function without interpolation in the 3D space, while the remaining two graphs present the projections of the values of the objective function on 2D plains (k-objective function and θs- objective function).

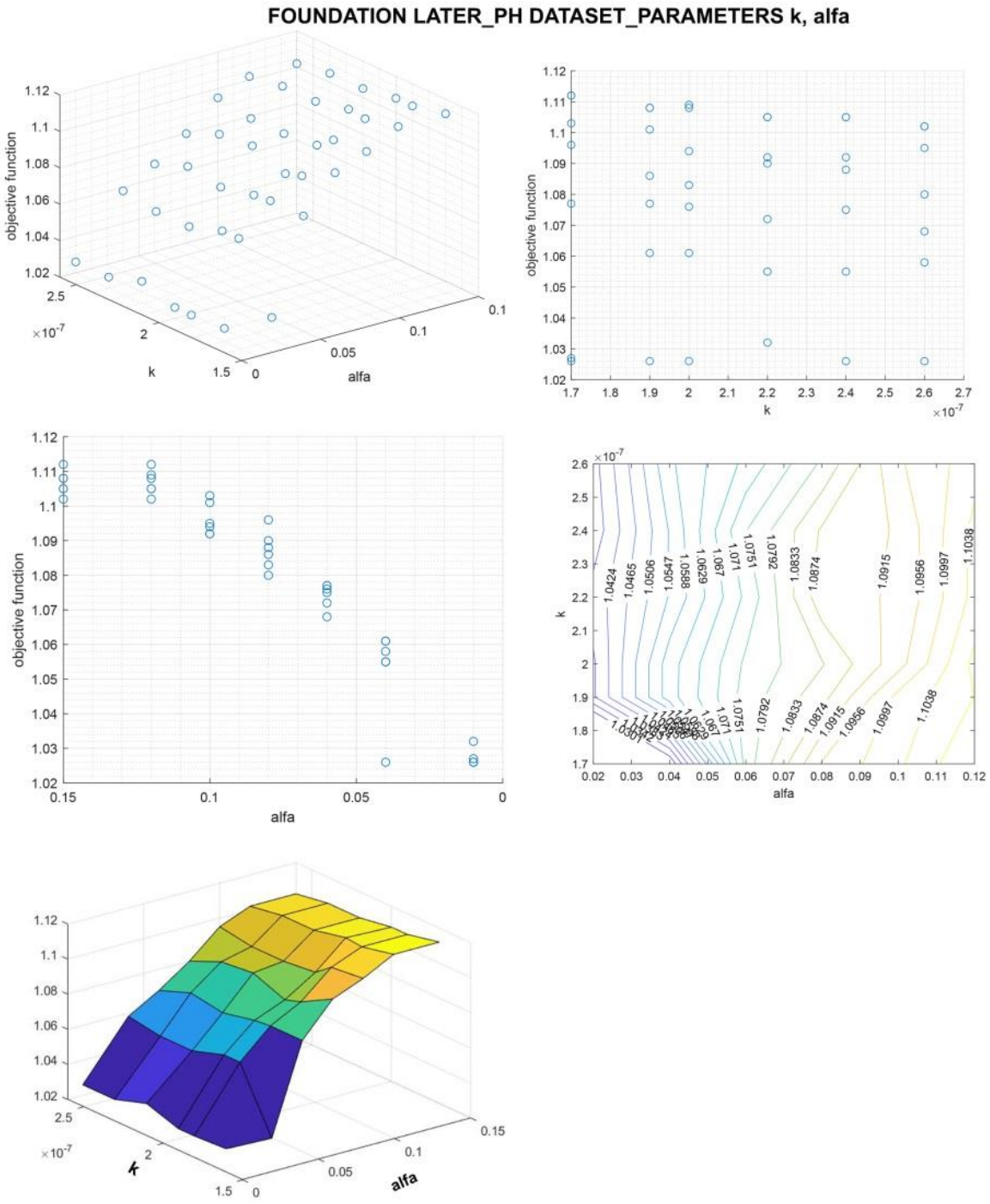


Figure 27:The behaviour of the objective function in the 2D/3D space of the optimized parameters k-α belonging to the foundation layer (HVGM) is presented. As it is possible to observe a clear minimum of the objective function has not been detected in the portion of the parameters space investigated.

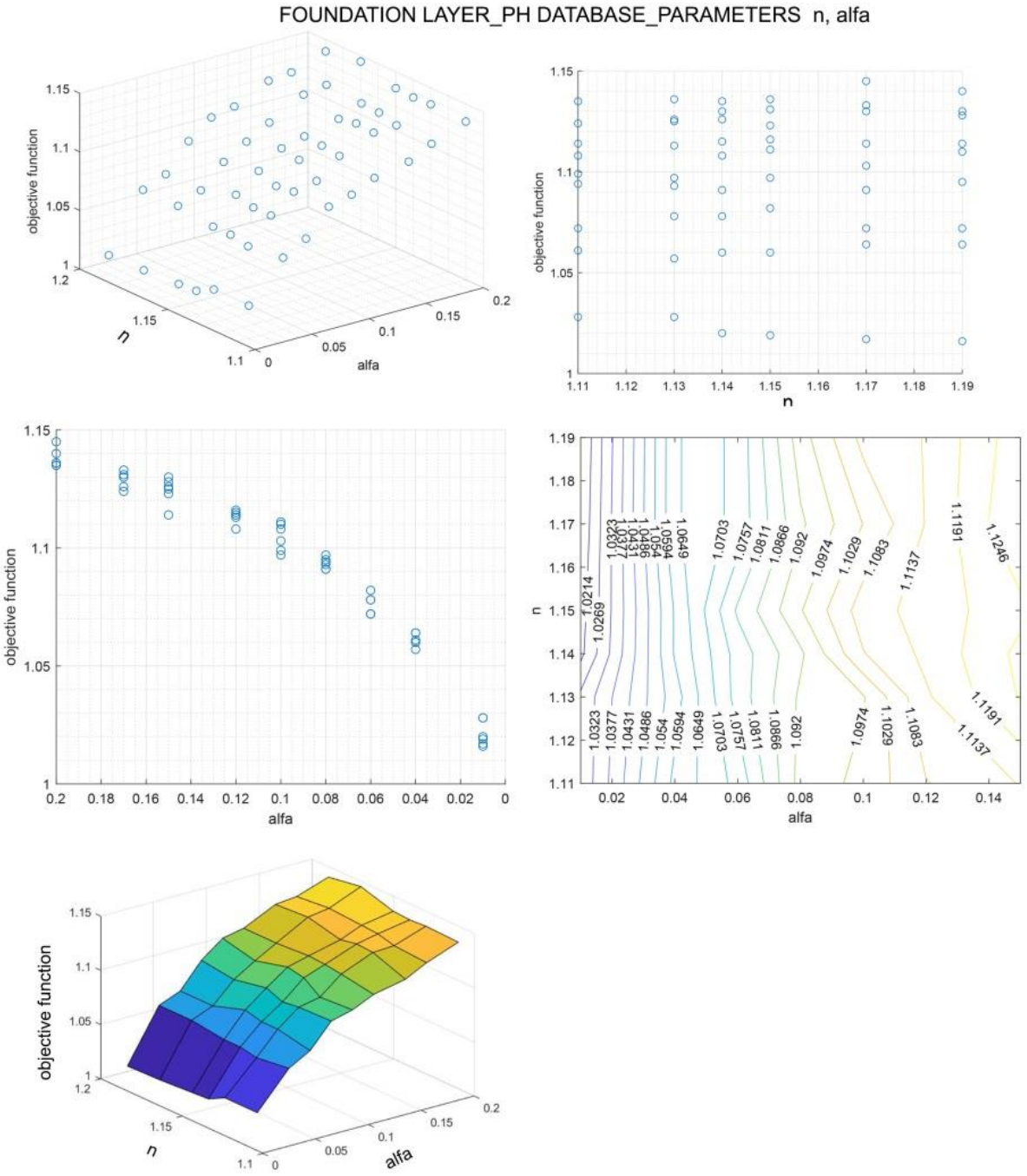


Figure 28: The behaviour of the objective function in the 2D/3D space of the optimized parameters n-α belonging to the foundation layer (HVGM) is presented. As it is possible to observe a clear minimum of the objective function has not been detected in the portion of the parameters space investigated.

7_4_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The indirect simulations that have overcome the first calibration phase are tested on a second phase in which the focus is on the optimized parameters and the goodness of them, looking closely at the last iteration of the inverse analysis. In this phase the modeler looks at the level of correlation between the optimized parameters, the tolerance to which the iteration process stops, the accuracy with which the finite difference matrix equations are being solved (Mass Balance Error), the number of optimized

parameters, the difference between the sum of the square residual at the zero iteration (using the initial set of parameters) and at the last iteration (using the optimized set of parameters), the number of iterations and the standard error coefficients of the optimized parameters. As in the first phase of the calibration, a scoring system is applied to each indirect simulations based on the distance of the obtained metrics values from the best fit. ''1'' is the maximum score associated to the best fit of each metric and 8 is the maximum score that each simulation could achieve. ''4'' is the lower acceptability score in order to pass to the next calibration phase. The'' quality parameters'' of the second phase are more stringent with respect to the ones of the first phase in order to decrease even more the number of indirect simulations.

53% of the initial number of simulations has overcome the second calibration phase and only 4 groups of inverse analysis have been excluded (zero successful simulations). The HVGM is the hydraulic model with the highest number of successful simulations, while VGM+2cm has the lowest percentage, due to an overall difficulty in the convergence of the indirect problem. The embankment layer has the highest number of successful indirect simulations (and among the highest scores) due to a greater number of simulations performed with respect to the other layers and a larger dataset of observation points used in the inverse problem. The embankment layer and the groups of layers (embankment+fluvial; embankment+foundation; embankment+foundation+fluvial) show the highest percentage of indirect simulations overcoming the second calibration phase, the only exception is the HVGM for the groups of layers embankment+foundation+fluvial. The VGM-2cm is excluded from the following calibration phase due to an insufficient performance of the inverse analysis. The poor performance of the indirect simulations that use the TMAT is confirmed also in the second phase of the calibration: we can conclude that the TMAT is not able to guarantee good results in the indirect problem. A small number of simulations that optimizes the hydraulic parameters of more than one layer has overcome both the calibration phases, this is a consequence of the greater difficulties and lower quality of the indirect simulations with increasing number of optimized parameters. Maintaining a low number of hydraulic parameters leads to a better performance of the inverse analysis in accordance with Simunek et al (2001) and Hopman et al (2002). Moreover it has been observed that a simulation using a pressure head dataset shows a score systematically lower (i.e. a poorer performance) with respect to the same simulation using a water content dataset. We can conclude that the use of a pressure head dataset allows to obtain better results in the indirect problem with respect to dataset of observation points of typology water content. With good probability this is due to the greater range of variability in which pwp measurements move in the 2018 simulation period with respect to the water content measurements (slow changes of small entity due to flooding events). Any simulation optimizing the parameters of the main wetting curve has overcome both the calibration phases (1^{st}+2^{nd}). This could be explained by the limited hysteretic effects on the riverbank behavior observed during the 2018 period which have been investigated extensively in chapter §6_1. Looking closely to the scores of each group of simulations important conclusions could be drawn: a low score attributed to the mass error indicates a low solution accuracy of the indirect problem; a low score associated to the ''number of iterations'' indicates that the obtained set of optimized parameters differs only slightly from the initial one. The high score attributed to the p-value suggests that the equations system is able to estimate singularly the optimized parameters while the high score in the simulation tolerance of the indirect simulations is proof of the high precision with which the inverse problem has been solved. It has been observed that an high score obtained in the first phase does not exclude problems of parameters non-identifiability, low accuracy in the parameters estimation, parameters correlations. The use of the Response Surface Analysis could be a valid help in case the behaviour of the objective function in proximity of a set of optimized parameters has to be investigated for problems of parameters correlation, ill-posedness of the inverse

problem, non-identifiability of a unique solution (multiple local minima) etc. This procedure is time consuming and for this reason, it is convenient to apply it only to a restricted number of cases (few couples of optimized parameters). The RSA could be used when the modeler looking at the values of the metrics/indices in the 1st and 2nd phases has detected problems in the inverse analysis and confirmations are required.

Chapter 7_6 investigates the influence of the initial set of hydraulic parameters on the results of the inverse analysis. The indirect simulations whose performances have been evaluated in the first and second calibration phases start from the same set of initial parameters (average values from laboratory tests). In chapter 7_6 a larger casuistry on possible initial sets of parameters has been applied to the indirect simulation of the 2018 period.

7_5 PERFORMANCE EVALUATION OF INDIRECT SIMULATIONS: THE THIRD CALIBRATION PHASE

After the second phase of the calibration, the remaining simulations have to be analysed in order to select the final combinations of parameters that enhance the most the performance of the investigated hydraulic models and whose performance will be tested on a new simulation period in the validation phase. It is extremely relevant to report that due to the fact that the optimized parameters have been obtained using different datasets of observation points (PH, wc, PH+wc) and optimizing different combinations of hydraulic parameters at a time belonging to one or more layers, a statistical analysis could not be considered properly representative of the results of the indirect simulations performed on the model. A statistical analysis is not robust enough to individuate the optimal parameter of an hydraulic model.
In fact an acceptable statistical sample has to be obtained under the same conditions in order to be representative of a population.

The number of performed inverse simulations is limited (~1100 simulations) with respect to a statistically significant number because the process of selection of the parameters to be optimized, the input dataset of observation points chosen and the weighting distributions given to each input observed data are manually set and not automatically performed by the software. For the same reason, it is not appropriate to refer to the combinations that show the better performances as the best combination of hydraulic parameters in absolute terms because there is not a statistical support behind it.
Briefly, not being able properly to use a probabilistic approach with which to suggest the final set of optimized parameters, a deterministic approach has been proposed in which a discrete number of simulations is analysed and the ones that show the best performance tested in the validation phase.

In the third phase of the calibration, the simulations to be investigated are the ones that have successfully undergone the second calibration phase. These simulations will be analysed using the same dataset of observation points, given by the available sensors which recorded data of suction and water content during the calibration period (2018) and using the same set of metrics (introduced in the first calibration phase) to quantify the fit between simulated-observed data. In the third phase of the calibration, the attention is focused on the performance of the single sensors and not on the overall performance of the whole set of available sensors as done in the first calibration phase. The set/s of hydraulic parameters that show also in this phase the best performance will be tested in the validation phase on the 2019 simulation period to investigate their predictive ability on a new period and on a new dataset of observation points.

In third calibration phase, the set of metrics chosen to evaluate the performance of the remaining 81 indirect simulations that have overcome the 2nd calibration phase are the following:

- -RMSE, R^2, MAE for the accuracy metrics
- -IA', NSE, NSE_j, NSE_{rel} for the efficiency metrics.

Robust and objective quantitative metrics that have shown a good ability to catch the behaviour of the model performance also in the first calibration phase. A lower number of metric has been adopted compared to the first phase of the calibration in order to maintain a manageable number of indices for each indirect simulation.

Once again, the attention is focused on the match observed-simulated data of the most relevant sensors which recorded in situ data during the calibration period. The considered 12 sensors are the following:
PZ1_10m (PH_foundation), MPS6_MPB2_2,7m (PH_foundation), PZ1_17m (PH_fluvial), T8_TB1_4,9m (PH_foundation), MPS6_MPC1_4,6m (PH_embankment); MPS6_MPC3_6,2m (PH_embankment); MPS6_SPC1_7,0m (PH_embankment); T8_TC2_8m (PH_embankment); GS3_MPB2_2,2m (w.c._fluvial); GS3_MPC1_2,4m (w.c._embankment); GS3_MPC1_4,5m (w.c._embankment) and GS3_SPC2_7,1m (w.c._embankment).

For each indirect simulation, the set of metrics has been applied to each sensor in order to investigate in detail its performance.
While the first calibration phase has focused the attention on the performance of all the available sensors together during the flood peak, off-peak periods and during the whole simulation period (peak+off-peak periods), in this phase the attention is concentrate on the performance of the single observation points/sensors during the whole simulation period and on the two main flood events of the 2018 period (flood C and G in `Figure 20`). Flood C has been recorded in the period 9/12/17-20/12/17 and flood G in the period 6/3/18-27/3/18. The flood events more persistent in time and that have reached a medium-high water level are the most relevant for a fine-grained riverbank as the one under investigation because the profile of suction has time to propagate more deeply in the bank body. The presented set of seven metrics is applied to the performance of each sensor on the whole simulation period and on the flooding periods C+G (n°12 sensors * 7 metrics *2 (floods events+whole period) = 168 metrics for each indirect simulation).

The reduction of the number of simulations (compared to the initial number) after calibration phases 1+2 allows to investigate their performance closer and in greater detail compared to the previous calibration phases. The final goal of the present calibration phase is obtaining the best possible match between in situ data and simulated data in the observation points that are of key importance for a good representation of the embankment behaviour during the simulated time period.

In order to ease the choice of the best indirect simulations, their performance is compared to the one of a reference simulation (base simulation) that uses the initial set of hydraulic parameters. The simulations that use HVGM will be compared to the performance of the base simulation that use the HVGM, while the simulations that use VGM will be compared to the base simulation VGM.
For each sensor, the difference Δ between the value of each metric of the investigated simulation and the value of the same metric of the simulation that uses the initial parameters dataset (base simulation) is calculated as follows [equation 143]:

$$\Delta_{ik} = v_{ik}^{opt} - v_{ik}^{init} \qquad \text{[equation 143]}$$

where i is the i-th sensor considered (i=1…12) j is the k-th metric considered (k=1…7)
v_{ik}^{opt} is the value of the metric k-th for the i-th sensor of the simulation to be investigated
v_{ik}^{init} is the value of the metric k-th for the i-th sensor of the base simulation.

The difference Δ_{ik} is positive for the metrics R^2 , IA', NSE, NSE_{rel} if the performance of the investigated simulation is better than the one of the base simulation because the best fit of these metrics is 1. The difference Δ_{ik} is negative for the metrics RMSE, MAE, NSE_j, if the performance of the investigated

simulation is better than the one of the base simulation because the best fit of these metrics is 0. For each sensor ''i'', the sum of the Δ_{ik} of the set of metrics is computed taking into account that the Δ_{ik} of RMSE, MAE and NSE_j is summed with a negative sign as follows [equation 144]:

$$\Delta_{i\_total} = \Delta_{i\_R^2\_} + \Delta_{i\_IA'} + \Delta_{i\_NSE} + \Delta_{i\_NSE_{rel}} - \Delta_{i\_RMSE} - \Delta_{i\_MAE} - \Delta_{i\_NSE_j} \quad \text{[equation 144]}$$

Due to the fact that the sensors are located in different positions in the bank section and they measure different physical properties (water potential or water content) they have a different ''weight'' in the model performance evaluation.

PZ1_10m and PZ2_17m are piezometers located at 10 m and 17 m from the soil surface in the saturated zone of the riverbank, their relevance for the representation of the pore pressure distribution during the investigated period is surely of minor importance compared to the sensors in the unsaturated zone of the section.

For this reason their weight is set equal to 0,5. The weight of the sensors which measure water content in the unsaturated zone as sensors GS3_MPB2_2,2m, GS3_MPC1_2,4m, GS3_MPC1_4,5m, GS3_SPC2_7,1m is set equal to 0,75 because the final aim of the modelling is the elaboration of a reliable pwp distribution in the most relevant instants of the simulation and the calculation is solved in terms of pressure head. The weights of the sensors which measure water potential in the unsaturated zone for the same reason are set equal to 1.

In a summary of the different ''weights'' for the considered sensors is reported:

Sensor	Weight	Sensor	Weight
PZ1_10m (PH)	0,5	MPS6_MPB2_2,7m (PH)	1
PZ2_17m (PH)	0,5	T8_TB1_4,9m (PH)	1
GS3_MPB2_2,2m (wc)	0,75	MPS6_MPC1_4,6m (PH)	1
GS3_MPC1_2,4m (wc)	0,75	MPS6_MPC3_6,2m (PH)	1
GS3_MPC1_4,5m (wc)	0,75	MPS6_SPC1_7,0m (PH)	1
GS3_SPC2_7,1m (wc)	0,75	T8_TC2_8m (PH)	1

Table 39 Weights attributed to each sensor according to their position in the bank section and the type of information that is recorded (water potential or water content)

For each sensor, Δ_{i_total} is multiplied by its weight (w_i). Then, for each simulation, we define Δ_{sim} :

$$\Delta_{sim} = \sum_{i=1}^{12} \Delta_{i\_total} \cdot w_i \quad \text{[equation 145]}$$

Δ_{sim} measures the global enhancement with respect to the base simulation.

Different weights are attributed to the set of chosen metrics/indices in order to have a balanced score between the accuracy metrics (RMSE, R^2, MAE) and efficiency metrics (IA', NSE, NSE_j, NSE_{rel}). In the different weights of the chosen metrics are listed:

ACCURACY METRICS		EFFICIENCY METRICS	
R^2	1	**IA'**	1
RMSE	0,5	**NSE**	0,33
MAE	0,5	**NSE_{rel}**	0,33
		NSE_j	0,33

Table 40 Weights attributed to the different accuracy and efficiency metrics

In and the values of Δ_{i_total} are reported for each considered sensor and for each indirect simulation that has overcome the second calibration phase (81 simulations) and in the last column the final score Δ_{sim} for each simulation is provided. reports the scores computed on the whole simulation period while the scores computed on the flood periods C+G. The indirect simulations that show the best performance on the two considered simulation periods are underlined in pink.

In the list of the indirect simulations that performed optimally during the whole simulation period and/or in the flood periods are reported: in red the 9 simulations with the highest score in both the investigated periods. These 9 indirect simulations will be tested in the validation phase on the 2019 simulation period.
In the scores (whole period) attributed to each sensor by the 9 simulations that perform better in the third calibration phase. In pink the sensors that show an enhancement with respect to the base simulation (score >0). As could be observed 3 sensors over 12 do not show an enhancement in any chosen indirect simulations (PZ1_10m; MPS6_SPC1_7m; GS3_MPC1_2,4m). From the values of the scores (>>0) it is possible to observe that the greatest enhancements are registered for GS3_MPC1_4,5m; MPS6_MPC1_4,6m; T8_TC2_8m and to a minor extent for GS3_SPC2_7,1m.
In the numbering used to refer to the considered sensors from here on out is reported.

FLOODS "C" AND "G" $\Delta_{i\_total}*w_i$													
	PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m	TOT SCORE Δsim
1_1_4	-41,8	-82,0	-41,2	-0,2	0,0	0,0	0,1	0,1	-0,1	-0,2	-0,2	-0,1	-165,8
1_1_5	153,8	300,6	151,1	0,9	0,0	0,0	-0,1	-0,1	0,2	0,5	0,5	0,2	607,6
1_1_15	-0,1	-0,3	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,6
1_1_9	2,2	4,3	2,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	8,6
1_1_11	85,3	167,0	83,9	0,5	0,0	0,0	-0,1	-0,1	0,1	0,3	0,3	0,1	337,4
1_1_12	-41,8	-82,0	-41,2	-0,2	0,0	0,0	0,1	0,1	-0,1	-0,2	-0,2	-0,1	-165,8
1_1_13	-6,6	-13,0	-6,5	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-26,2
1_1_14	23,7	46,5	23,4	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	93,7
1_1_16	-0,3	-0,6	-0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,1
1_1_17	143,6	280,6	141,1	0,8	0,0	0,0	-0,1	-0,1	0,2	0,5	0,4	0,2	567,2
1_2_2	342,0	657,5	331,7	3,3	0,0	-0,1	-0,4	-0,3	0,6	1,9	1,7	0,6	1338,4
1_2_6	339,6	654,4	330,1	3,1	-0,3	-0,4	-0,5	-0,5	0,0	1,2	1,2	-0,1	1327,8
1_2_12	336,6	649,2	327,3	3,0	-0,1	-0,2	-0,3	-0,3	0,4	1,7	1,6	0,4	1319,3
1_3_5	184,3	359,8	180,9	1,1	0,0	0,0	-0,2	-0,2	0,3	0,7	0,6	0,3	727,6
1_3_12	30,9	60,5	30,4	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	121,9
1_3_15	-0,1	-0,3	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,6
1_3_9	-3,4	-6,7	-3,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-13,6
1_3_10	-21,0	-41,1	-20,7	-0,1	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1	-0,1	-83,1
1_3_11	85,3	167,0	83,9	0,5	0,0	0,0	-0,1	-0,1	0,1	0,3	0,3	0,1	337,4
1_3_13	-72,4	-142,0	-71,3	-0,3	0,0	0,0	0,1	0,1	-0,1	-0,4	-0,3	-0,2	-286,8
1_3_14	23,7	46,5	23,4	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	93,7
1_3_16	-0,3	-0,6	-0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,1
1_3_17	99,6	194,9	97,9	0,5	0,0	0,0	-0,1	-0,1	0,1	0,2	0,2	0,1	393,5
1_4_3	337,8	649,9	327,7	3,2	0,2	0,1	-0,4	-0,2	1,1	2,4	2,1	1,3	1325,1
1_4_6	335,9	648,3	327,0	2,9	-0,3	-0,4	-0,4	-0,5	-0,2	0,7	0,8	-0,1	1313,7
1_4_7	323,1	625,2	315,1	2,6	-0,2	-0,4	-0,5	-0,5	0,0	1,1	1,1	-0,1	1266,5
1_7_10	-0,4	-0,8	-0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,7
1_9_3	-0,3	-0,6	-0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,2
1_9_10	0,1	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,6
1_9_13	-0,6	-1,1	-0,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-2,3
1_9_16	-3,3	-6,4	-3,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-13,0
1_11_3	341,9	657,4	331,7	3,3	-0,1	-0,2	-0,4	-0,3	0,3	1,6	1,5	0,3	1337,1
1_11_4	334,8	645,3	325,2	3,0	0,2	0,1	-0,4	-0,2	1,0	2,4	2,1	1,3	1314,9
1_13_3	14,0	27,5	13,8	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	55,6
1_13_4	317,9	619,9	311,8	2,1	0,1	0,0	0,2	0,2	-0,2	-0,5	-0,4	-0,3	1250,8
1_13_10	-40,4	-79,2	-39,8	-0,2	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1	0,0	-159,7
1_13_11	-19,2	-37,6	-18,9	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-75,8
1_13_13	-16,9	-33,1	-16,6	-0,1	0,0	0,0	0,0	0,0	0,0	-0,1	-0,1	0,0	-67,0
1_13_15	-0,7	-1,4	-0,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-2,9
1_15_15	1,6	3,2	1,6	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	6,5
1_17_14	2,2	4,3	2,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	8,6
1_17_16	2,2	4,3	2,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	8,6
1_39_3	-0,6	-1,1	-0,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-2,3
1_10_12	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,00
1_10_14	-0,3	-0,6	-0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,11
1_10_10	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,00
1_10_17	0,6	1,1	0,6	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,31
1_14_2	-1,2	-2,4	-1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	-4,62
1_14_3	64,2	125,7	63,1	0,3	0,0	0,0	-0,1	-0,1	0,1	0,4	0,3	0,2	254,18
1_14_4	15,2	29,7	14,9	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	60,09
1_14_11	-15,5	-30,3	-15,2	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-61,16
1_14_15	1,3	2,5	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,0	0,0	5,17
1_12_4	337,9	650,1	327,8	3,2	0,2	0,1	-0,4	-0,2	1,1	2,4	2,1	1,3	1325,60
1_16_2	-0,1	-0,3	-0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-0,58
1_16_4	10,6	20,8	10,5	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	42,15
1_16_5	3,2	6,2	3,1	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	12,65
1_16_6	-0,4	-0,9	-0,4	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-1,77
1_16_10	1,9	3,7	1,9	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	7,51
1_16_13	4,4	8,6	4,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	17,52
1_16_15	-1,3	-2,6	-1,3	0,0	0,0	0,0	-0,1	0,0	0,1	0,1	0,1	0,1	-5,00
1_16_17	2,3	4,5	2,3	0,0	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	9,28
1_18_10	0,3	0,6	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,11
1_18_11	0,1	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,55
1_18_14	0,1	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,56
1_18_15	0,3	0,6	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,11
1_18_16	0,1	0,3	0,1	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,56
1_18_17	0,3	0,6	0,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,11
1_56_10	-15893,1	-31618,1	-15829,3	-20,9	-0,5	2,6	29,6	47,0	-24,9	-81,4	-57,6	-31,9	-6,35E+04
1_5_3	4,1	8,0	4,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	16,20
1_5_4	-11,3	-22,1	-11,1	-0,1	0,0	0,0	0,0	0,0	0,0	-0,1	0,0	0,0	-44,63
1_5_5	66,5	130,1	65,4	0,4	0,0	0,0	-0,1	-0,1	0,1	0,2	0,2	0,1	262,86
1_5_11	22,3	43,7	22,0	0,1	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	88,26
1_5_12	-5,3	-10,3	-5,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-20,85
1_5_14	29,4	57,6	28,9	0,2	0,0	0,0	0,0	0,0	0,0	0,1	0,1	0,0	116,25
1_5_10	-30,1	-59,1	-29,7	-0,1	0,0	0,0	0,1	0,1	-0,1	-0,2	-0,2	-0,1	-119,44
1_5_13	-0,7	-1,4	-0,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	-2,91
1_5_16	0,7	1,4	0,7	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	2,83
1_5_17	42,2	82,7	41,6	0,2	0,0	0,0	0,0	0,0	0,1	0,2	0,1	0,1	167,12
1_6_3	342,1	657,7	331,8	3,3	0,0	-0,1	-0,4	-0,3	0,6	2,0	1,8	0,7	1338,96
1_6_9	340,2	655,4	330,7	3,1	-0,2	-0,4	-0,5	-0,5	0,1	1,3	1,2	0,0	1330,45

Table 41: Scores $\Delta_{i\_total}*w_i$ attributed to each considered sensor of each indirect simulation that has overcome the second calibration phase. Metrics are applied to the flood period (C+G).

WHOLE SIMULATION PERIOD Δi_total*wi													
	PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m	TOT SCORE Δ_{sim}
1_1_4	0,01	0,07	0,15	-0,03	3,39	-0,14	0,04	8,11	-0,05	4,73	-3,47	0,18	12,98
1_1_5	0,00	0,12	0,12	-0,03	-21,79	-0,13	0,05	7,31	-0,04	-55,18	223,56	0,51	154,52
1_1_15	0,01	0,03	0,12	-0,02	0,68	-0,01	0,00	-0,22	-0,02	0,04	-0,36	0,01	0,25
1_1_9	0,00	0,00	0,02	0,00	4,29	0,01	0,00	-0,09	0,01	0,50	5,50	0,02	10,26
1_1_11	0,00	0,06	0,01	0,00	-12,81	-0,07	0,04	4,73	0,00	-18,85	124,03	0,28	97,43
1_1_12	0,00	0,01	0,01	0,01	0,96	0,00	0,01	0,69	0,01	0,10	-61,47	-0,28	-59,96
1_1_13	0,00	0,00	0,01	0,00	3,31	0,00	0,01	0,42	0,01	0,86	-8,57	-0,04	-3,99
1_1_14	0,00	0,05	0,03	0,00	-8,96	-0,06	0,04	4,32	0,00	-5,33	33,78	-0,05	23,84
1_1_16	0,00	0,00	0,02	0,00	0,25	0,00	0,00	0,05	0,01	-0,03	-0,13	0,00	0,17
1_1_17	0,00	0,12	0,02	-0,01	-21,23	-0,16	0,05	9,86	0,00	-48,23	208,97	0,44	149,84
1_2_2	-0,19	0,02	0,16	-0,16	231,19	-0,88	-1,00	28,18	0,45	-139,96	538,39	2,96	659,14
1_2_6	-0,36	0,16	0,04	-0,11	230,61	-1,95	-1,68	44,02	0,43	-91,19	535,14	2,67	717,79
1_2_12	-0,22	-0,04	0,05	-0,15	236,08	-0,94	-1,08	29,77	0,57	-188,07	532,65	3,07	611,68
1_3_5	0,00	0,15	0,13	-0,03	-28,06	-0,17	0,05	9,48	-0,03	-81,40	269,62	0,65	170,40
1_3_12	0,01	0,10	0,15	-0,03	-17,97	-0,09	0,04	4,77	-0,04	-9,69	40,90	-0,11	18,05
1_3_15	0,01	0,03	0,12	-0,02	0,68	-0,01	0,00	-0,22	-0,02	0,04	-0,36	0,01	0,25
1_3_9	0,00	0,00	0,06	0,01	-4,39	0,00	-0,01	-0,26	0,00	-0,69	-6,18	-0,02	-11,47
1_3_10	0,00	0,01	0,02	0,00	0,38	0,00	0,01	0,50	0,01	0,55	-31,23	-0,13	-29,90
1_3_11	0,00	0,06	0,01	0,00	-12,81	-0,07	0,04	4,73	0,00	-18,85	124,03	0,28	97,43
1_3_13	0,00	0,02	0,03	0,01	-3,69	-0,02	0,02	1,42	0,00	-2,05	-104,81	-0,44	-109,53
1_3_14	0,00	0,05	0,03	0,00	-8,96	-0,06	0,04	4,32	0,00	-5,33	33,78	-0,05	23,84
1_3_16	0,00	0,00	0,02	0,00	0,25	0,00	0,00	0,05	0,01	-0,03	-0,13	0,00	0,17
1_3_17	0,01	0,10	0,07	0,01	-29,96	-0,12	0,05	7,22	-0,01	-32,61	141,67	0,16	86,59
1_4_3	-0,02	-0,15	0,02	0,03	-16,78	0,03	-0,27	-12,86	-0,04	-376,77	530,17	3,20	126,56
1_4_6	-0,49	0,24	0,07	-0,11	234,95	-2,78	-2,47	58,70	0,42	-58,70	530,28	2,49	762,59
1_4_7	-0,36	0,23	0,10	0,04	37,66	-3,08	-2,17	58,44	-0,27	-123,37	511,02	2,81	481,05
1_7_10	0,00	-0,04	0,03	-0,11	1,73	0,03	0,02	-0,35	26,43	-0,07	-0,62	-0,02	27,02
1_9_3	0,00	0,02	0,12	-0,03	1,78	0,00	-0,01	-0,47	-0,02	0,03	-0,68	0,00	0,74
1_9_10	0,00	0,02	0,04	0,01	-1,65	-0,01	0,00	0,61	-0,01	-0,12	0,63	0,01	-0,47
1_9_13	0,00	0,01	0,01	0,02	1,16	0,01	0,00	-0,07	-0,02	-0,08	-0,40	-0,01	0,62
1_9_16	0,00	0,00	0,06	0,01	-3,21	0,00	-0,01	-0,34	-0,02	-0,78	-5,85	-0,03	-10,16
1_11_3	-0,24	0,00	0,06	0,38	228,54	-1,30	-1,27	34,62	3,24	-142,10	538,75	3,01	663,71
1_11_4	-0,02	-0,24	0,01	0,01	-16,22	0,04	-0,24	-12,03	26,38	-335,37	521,50	3,07	186,88
1_13_3	0,01	0,03	0,14	-0,01	13,84	-0,01	0,01	0,12	-0,01	2,03	21,95	0,09	38,19
1_13_4	0,01	0,04	0,15	-0,02	2,97	-0,03	0,02	1,54	-0,02	-4,90	-154,75	-0,66	-155,66
1_13_10	0,00	0,04	0,06	0,04	-36,73	-0,08	0,01	3,71	-0,11	-3,53	-60,70	-0,13	-97,42
1_13_11	0,01	0,03	0,10	0,01	-1,10	-0,07	0,03	4,30	-0,02	2,02	-26,07	-0,03	-20,78
1_13_13	0,00	0,01	0,00	0,00	-1,71	0,00	0,00	0,27	0,00	0,20	-24,25	-0,11	-25,59
1_13_15	0,00	0,01	0,02	0,00	0,58	0,00	0,01	0,66	0,00	0,29	-0,60	0,01	0,98
1_15_15	0,01	0,02	0,24	0,03	-2,82	-0,06	-0,01	1,51	-0,24	-0,05	3,25	0,05	1,94
1_17_14	0,01	0,02	0,16	-0,02	5,42	0,00	0,00	-0,44	-0,02	0,59	5,11	0,02	10,84
1_17_16	0,01	0,02	0,16	-0,02	5,42	0,00	0,00	-0,44	-0,02	0,59	5,11	0,02	10,84
1_39_3	0,00	-0,07	0,12	-0,05	2,66	0,01	0,00	-0,58	26,40	0,05	-0,95	-0,01	27,58
1_10_12	0,00	0,01	0,10	-0,05	0,27	0,00	0,00	-0,09	0,01	-0,02	-0,05	0,00	0,18
1_10_14	0,00	-0,01	-0,04	-0,08	0,83	0,01	0,00	-0,07	0,04	-0,01	-0,27	-0,01	0,40
1_10_10	0,00	-0,01	0,00	-0,05	0,16	0,00	0,00	-0,11	0,01	-0,02	-0,03	0,00	-0,05
1_10_17	0,01	0,04	-0,02	-0,02	-3,09	-0,02	0,01	1,24	0,00	-0,04	1,23	0,02	-0,64
1_14_2	0,01	0,02	0,08	-0,03	-0,49	-0,04	0,02	2,68	-0,02	1,60	-0,64	0,06	3,26
1_14_3	0,00	-0,01	0,11	-0,03	56,37	0,01	0,04	1,31	0,10	7,32	103,06	0,38	168,65
1_14_4	0,01	0,00	0,05	-0,03	-0,83	0,00	0,00	-0,31	0,00	-0,71	21,44	0,10	19,70
1_14_11	0,02	0,05	0,03	0,08	-4,23	-0,03	0,01	1,51	0,00	0,36	-20,14	-0,07	-22,42
1_14_15	0,01	0,05	-0,07	-0,03	-6,39	-0,04	0,01	1,66	-0,02	-0,14	2,37	0,05	-2,53
1_12_4	-0,02	-0,25	0,04	-0,04	-15,99	0,04	-0,26	-13,18	26,35	-375,45	529,82	3,20	154,26
1_16_2	0,01	-0,01	0,00	-0,05	0,42	0,01	-0,01	-0,48	0,01	-0,21	-0,02	-0,01	-0,34
1_16_4	0,01	0,02	-0,06	-0,06	0,83	-0,01	0,01	0,85	5,71	0,52	16,89	0,09	24,80
1_16_5	0,01	0,05	-0,03	0,01	-2,11	-0,03	0,02	1,87	-0,03	0,45	7,58	0,06	7,85
1_16_6	0,01	0,00	0,08	-0,04	-0,66	0,00	0,00	-0,04	0,00	0,07	-0,73	0,00	-1,32
1_16_10	0,01	0,04	-0,12	-0,04	-1,36	-0,01	0,00	0,33	0,05	-0,29	3,63	0,01	2,25
1_16_13	0,00	0,00	0,03	-0,05	-1,28	0,00	0,00	-0,24	-0,09	-0,24	8,24	0,04	6,42
1_16_15	0,01	0,09	-0,02	-0,05	-3,52	-0,08	0,04	5,05	0,00	2,10	0,32	0,10	4,02
1_16_17	0,01	0,04	-0,01	-0,04	-5,39	-0,02	0,01	1,31	0,08	-0,31	4,88	0,04	0,59
1_18_10	0,01	0,01	0,05	-0,02	-0,89	0,00	0,00	-0,06	0,00	-0,03	0,27	0,00	-0,68
1_18_11	0,01	0,00	0,00	-0,05	-0,40	0,00	0,00	-0,23	0,01	-0,05	0,19	0,00	-0,52
1_18_14	0,01	0,00	0,04	-0,05	-0,59	0,00	0,00	-0,21	0,01	-0,02	0,22	0,00	-0,61
1_18_15	0,00	0,00	0,01	-0,05	-0,35	0,00	0,00	-0,32	0,01	-0,02	0,22	0,00	-0,50
1_18_16	0,00	0,00	0,00	-0,04	-0,34	0,00	0,00	-0,09	0,00	-0,01	0,11	0,00	-0,36
1_18_17	0,01	0,01	0,09	-0,04	-0,75	0,00	0,00	-0,19	0,00	-0,01	0,27	0,00	-0,61
1_56_10	-2,20	6,91	-34,35	-0,62	-1,07E+04	-11,96	-6,89	69,65	-4,87E+02	-1,98E+04	-1,94E+04	-5,04E+01	-5,04E+04
1_5_3	0,01	-0,01	0,02	-0,05	5,72	0,01	0,01	0,06	0,02	0,88	9,27	0,03	15,97
1_5_4	0,00	0,00	0,10	-0,04	-0,47	0,00	0,00	0,16	0,01	0,29	-15,70	-0,07	-15,70
1_5_5	0,00	0,05	0,06	-0,04	-10,50	-0,05	0,03	3,14	0,00	-12,37	95,27	0,20	75,81
1_5_11	0,01	0,02	0,04	-0,04	-4,69	-0,02	0,01	0,87	0,00	-3,09	33,19	0,07	26,38
1_5_12	0,01	0,00	0,04	-0,05	6,58	-0,01	0,01	0,30	0,02	1,38	-6,99	-0,03	1,28
1_5_14	0,01	0,01	0,07	-0,05	1,36	-0,02	0,02	1,28	0,02	-2,83	43,41	0,08	43,37
1_5_10	0,00	0,01	0,06	-0,04	0,62	0,00	0,00	-0,04	0,01	-0,01	-43,57	-0,21	-43,17
1_5_13	0,00	-0,01	0,06	-0,04	5,66	0,00	0,00	-0,14	0,02	0,86	-0,58	-0,02	5,83
1_5_16	0,01	-1,26E+05	-1,05E+05	-5,19E+04	-1,06E+05	-1,06E+05	-1,07E+05	-1,49E+05	-8,01E+04	-8,01E+04	-7,96E+04	-8,01E+04	-1,07E+06
1_5_17	0,01	0,02	0,05	-0,04	0,51	-0,02	0,02	1,52	0,02	-5,09	62,57	0,13	59,70
1_6_3	-0,18	-0,12	-0,10	-0,23	232,31	-0,80	-0,94	27,58	0,65	-143,57	538,44	2,96	656,00
1_6_9	-0,32	0,04	-0,05	-0,22	231,73	-1,71	-1,53	40,28	0,60	-97,01	535,64	2,72	710,16

Table 42: Scores $\Delta_{i\_total}*w_i$ attributed to each considered sensor of each indirect simulation that has overcome the second calibration phase. Metrics are applied to the whole simulation period.

SCORE $\Delta_{i\_total}*w_i$		FLOODS "C" AND "G" Δi_total*wi	WHOLE PERIOD Δi_total*wi
1_2_2	HVGM_wc_embankment_α	1338,4	659,1
1_2_6	HVGM_wc_embankment_k,α	1327,8	717,8
1_2_12	HVGM_wc_embankment_k,α,θr	1319,3	611,7
1_4_3	HVGM_wc_embankment_θs	1325,1	126,6
1_4_6	HVGM_wc_embankment_k,α	1313,7	762,6
1_4_7	HVGM_wc_embankment_k,θs	1266,5	481,1
1_11_3	VGM_wc_embank+fluvial_α	1337,1	663,7
1_11_4	VGM_wc_embank+fluvial_θs	1314,9	186,9
1_13_4	HVGM_PH_embank+found_θs	1250,8	-155,7
1_12_4	VGM_wc_embank+fluv_θs	1325,6	154,3
1_6_3	VGM_wc_embankment_α	1339,0	656,0
1_6_9	VGM_wc_embankment_α,k	1330,5	710,2

Table 43: List of indirect simulations that have shown the best performance in the third phase of the calibration. In red the simulations that have obtained a good score both in the performance during the flood periods (C+G) and during the whole simulation period.

SCORE (WHOLE SIMULATION PERIOD) $\Delta_{i\ total}*w_i$		PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m	TOT SCORE
1_2_2	HVGM_wc_embankment_α	-0,19	0,02	0,16	-0,16	231,19	-0,88	-1,00	28,18	0,45	-139,96	538,39	2,96	659,14
1_2_6	HVGM_wc_embankment_k,α	-0,36	0,16	0,04	-0,11	230,61	-1,95	-1,68	44,02	0,43	-91,19	535,14	2,67	717,79
1_2_12	HVGM_wc_embankment_k,α,θr	-0,22	-0,04	0,05	-0,15	236,08	-0,94	-1,08	29,77	0,57	-188,07	532,65	3,07	611,68
1_4_6	HVGM_wc_embankment_k,α	-0,49	0,24	0,07	-0,11	234,95	-2,78	-2,47	58,70	0,42	-58,70	530,28	2,49	762,59
1_4_7	HVGM_wc_embankment_k,θs	-0,36	0,23	0,10	0,04	37,66	-3,08	-2,17	58,44	-0,27	-123,37	511,02	2,81	481,05
1_11_3	VGM_wc_embank+fluvial_α	-0,24	0,00	0,06	0,38	228,54	-1,30	-1,27	34,62	3,24	-142,10	538,75	3,01	663,71
1_11_4	VGM_wc_embank+fluvial_θs	-0,02	-0,24	0,01	0,01	-16,22	0,04	-0,24	-12,03	26,38	-335,37	521,50	3,07	186,88
1_6_3	VGM_wc_embankment_α	-0,18	-0,12	-0,10	-0,23	232,31	-0,80	-0,94	27,58	0,65	-143,57	538,44	2,96	656,00
1_6_9	VGM_wc_embankment_α,k	-0,32	0,04	-0,05	-0,22	231,73	-1,71	-1,53	40,28	0,60	-97,01	535,64	2,72	710,16

Table 44: Scores (whole period) attributed to each sensor by the 9 simulations that perform best in the third calibration phase. In pink the sensors that show an enhancement with respect to the base simulation

	PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m
Sensor n°	1	2	3	4	5	6	7	8	9	10	11	12

Table 45: The table reports the numbering of the considered sensors that will be used from here on out.

Summarising, 9 indirect simulations have shown the greatest enhancements with respect to the base simulation in the third phase of calibration: 5 simulations for the HVGM and 4 for the VGM:

1_2_2: HVGM_enhancement of the performance of sensors 2, 3, 5, 8, 9, 11, 12 worsening of the performance of sensors 1, 4, 6, 7,10.

1_2_6: HVGM_enhancement of the performance of sensors 2, 3, 5, 8, 9, 11, 12, worsening of the performance of sensors 1, 4, 6, 7,10.

1_2_12: HVGM_enhancement of the performance of sensors 3, 5, 8, 9, 11, 12; worsening of the performance of sensors 1, 2, 4, 6, 7, 10.

1_6_3: VGM_enhancement of the performance of sensors 5, 8, 9, 11, 12, worsening of the performance of the sensors 1, 2, 3, 4, 6, 7, 10.

1_6_9: VGM_enhancement of the performance of sensors 2, 5, 8, 9, 11, 12; worsening of the performance of sensors 1, 3, 4, 6, 7, 10.

1_11_3: VGM_enhancement of the performance of sensors 2, 3, 4, 5, 8, 9, 11, 12; worsening of the performance of sensors 1, 6, 7, 10.

1_11_4: VGM_enhancement of the performance of sensors 3, 4, 6, 8, 9, 11, 12; worsening of the performance of sensors 1, 2, 5, 7, 10.
1_4_6: HVGM_enhancement of the performance of sensors 2, 3, 5, 8, 9, 11, 12 ; worsening of the performance of sensors 1, 4, 6, 7, 10.
1_4_7: HVGM_enhancement of the performance of the sensors 2, 3, 4, 5, 8, 11, 12; worsening of the performance of sensors 1, 6, 7, 9, 10.

As it is possible to notice the indirect simulations that have overcome the third calibration phase optimize the hydraulic parameters of the embankment layer (7/9) and the parameters of embankmet+fluvial (2/9).
The whole set of 9 indirect simulation uses a wc dataset and a weighting distribution w=1.

It is important to remember that a negative value of $\Delta_{i\_total} \cdot w_i$ means a worsening of the performance of the sensor compared to the base simulation (that uses the initial set of hydraulic parameters) while a positive value denotes an enhancement of the performance of the considered sensor. It must be also emphasized that values Δ_{i_tot} below +5 and above -5 are almost impossible to detect in the graphical representation of the sensors trend in time when compared to the base simulation.

For this reason, only sensors 5, 8, 10, 11 show enhancements of considerably importance in . Simulation 1_11_3 has the highest number of sensors whose performances have been enhanced (8/12 sensors) compared to the base simulation and the lower number of sensors whose performance is reduced (4/12). Indirect simulations 1_2_2, 1_2_6, 1_4_6 optimize the same sensors but the best performance (higher score) is shown by simulation 1_4_6.
As could be observed, the majority of the simulations enhances the performance of some sensors and reduces the performance of others, no simulation is able to enhance the performance of all the sensors together, for this reason it is relevant to focus the attention on those observation points that are most significant for the analysis.
Sensors of water potential positioned in the unsaturated zone of the embankment layer such as sensors 2 (MPS6 MPB2 2,7m), 5 (MPS6 MPC1 4,6m), 6 (MPS6-MPC3 6,2m), 7 (MPS6 SPC1 7m) and sensor 4 (T8 TB1 4,7m) positioned in the unsaturated zone of the berm are the most significant to elaborate an accurate map of pore water pressure for future stability analysis.

To the set of 9 indirect simulations that has overcome the third calibration phase other 2 simulations are added. These additional simulations (2_1_3 and 2_3_3) are performed in chapter §7_7 to investigate the role of different weighting distributions on the results of the inverse analysis. The third calibration phase is applied also to these additional simulations (2_1_3 and 2_3_3). The comparison showed that 2_1_3 and 2_3_3 were the best performing in chapter §7_7 (Table 43). In this case the indirect simulations use a dataset of type PH.

In the scores for the whole simulation period and for the two flood periods (C+G) are presented for the new indirect simulations (2_1_3 and 2_3_3). As could be observed both the simulations enhance the performance of sensors 5, 8, 9, 11, 12 for the whole simulation period and of sensors 1, 2, 3, 4, 5, 9, 10, 11, 12 for the flood periods (C+G).

It is relevant to point out that changes in the sensors performance are in all cases of small entity due to the fact that the set of initial parameters is obtained from a reliable laboratory campaign. Reliable laboratory tests allow starting from a good performance of the model and save time in the calibration phase, avoiding a high number of inverse analyses to perform especially when a multi-layered structure is investigated. The initial parameters chosen in the presented case study showed to be extremely accurate and for this reason the optimized parameters obtained from inverse analysis are located in a narrow interval close to the initial parameters.

This could be easily observed in where the sets of optimized parameters of the 11 chosen indirect simulations are presented (in red the values that show changes with respect to the values of the initial parameters).

SCORE (WHOLE SIMULATION PERIOD) $\Delta_{i\_total}*w_i$														
		PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m	TOT SCORE
2_1_3	VGM_emb_PH_flood periods=1; off-flood periods=0; α.	-0,1	-0,1	0,0	-0,1	183,0	-0,3	-0,3	13,7	0,4	-8,9	413,0	1,9	602,4
2_3_3	VGM_emb_PH_flood periods=weights dependent of water h and permanence; off-flood periods=0; α.	0,0	-0,1	-0,1	-0,1	178,9	-0,3	-0,2	14,2	0,4	-6,0	402,4	1,8	590,9
SCORE (FLOODS "C" AND "G") $\Delta_{i\_total}*w_i$														
2_1_3	VGM_emb_PH_flood periods=1; off-flood periods=0; α.	260,1	506,1	254,5	1,7	0,1	-0,1	-0,4	-0,4	0,6	1,7	1,5	0,7	1026,3
2_3_3	VGM_emb_PH_flood periods=weights dependent of water h and permanence; off-flood periods=0; α.	253,6	493,6	248,2	1,7	0,1	-0,1	-0,4	-0,4	0,6	1,7	1,5	0,7	1000,8

Table 46: Scores attributed in the third calibration phase to the new indirect simulations (see chapter §7_7). In pink the sensors that show an enhancement ($\Delta_{i\_total}*w_i>0$) with respect to the performance of the base simulation.

	embankment									fluvial								
	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw
1_2_2_HVGM	0,004	0,397	1,704	1,304	9,00E-05	0,398	0,397	3,408	9,00E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_2_6_HVGM	0,004	0,397	1,640	1,304	7,49E-05	0,398	0,397	3,280	7,49E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_2_12_HVGM	0,044	0,397	1,830	1,304	7,89E-05	0,398	0,397	3,660	7,89E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_4_6_HVGM	0,004	0,397	1,609	1,304	6,26E-05	0,398	0,397	3,218	6,26E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_4_7_HVGM	0,004	0,351	0,850	1,304	4,31E-05	0,352	0,351	1,700	4,31E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_11_3_HVGM	0,004	0,397	1,708	1,304	9,00E-05	0,398	0,397	3,416	9,00E-05	0,0005	0,319	0,027	1,284	2,76E-05	0,319	0,319	0,053	2,76E-05
1_11_4_HVGM	0,004	0,351	0,850	1,304	9,00E-05	0,351	0,351	1,700	9,00E-05	0,0005	0,399	0,251	1,284	2,76E-05	0,399	0,399	0,502	2,76E-05
1_6_3_VGM	0,004	0,397	1,710	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
1_6_9_VGM	0,004	0,397	1,647	1,304	7,68E-05					0,0005	0,319	0,251	1,284	2,76E-05				
2_1_3_VGM	0,004	0,397	1,239	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
2_3_3_VGM	0,004	0,397	1,223	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
initial paramete	0,004	0,397	0,850	1,304	9,00E-05	0,398	0,397	1,700	9,00E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
	foundation									aquifer								
	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw
1_2_2_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_2_6_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_2_12_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_4_6_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_4_7_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_11_3_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_11_4_HVGM	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_6_3_VGM	0,001	0,424	0,118	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
1_6_9_VGM	0,001	0,424	0,118	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
2_1_3_VGM	0,001	0,424	0,118	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
2_3_3_VGM	0,001	0,424	0,118	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
initial paramete	0,001	0,424	0,118	1,142	1,98E-07	0,424	0,424	0,236	1,980E-07	0,010	0,43	0,2	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05

Table 47: Sets of optimized parameters that have overcome the third calibration phase (11 indirect simulations). In red the values of the hydraulic parameters changed with respect to the set of initial parameters (last row). Ks and Ksw are in m/min (saturated permeability of the drying and wetting curve respectively), α and αw in 1/m, θs, θr, θm and θsw have no physical dimensions.

In and , the performances of the two indirect simulations (1_11_4; 2_1_3) chosen among the 11 available are compared with the base simulation and the observed dataset for each considered sensor in order to visually observe the behaviour highlighted by the chosen set of metrics/indices in . Indirect simulation 1_11_4 has used a wc dataset to perform the inverse analysis while 2_1_3 a PH dataset.
As could be observed in both the graphs, sensors monitoring the water content using the set of optimized parameters seem to perform better with respect to sensors monitoring the water potential. This is due to the fact that the base simulation shows a very good performance in the simulation of the pwp distribution

(with metrics close to the best fit) while it shows a poorer performance in the simulation of the water content distribution. For this reason, obtaining an enhancement in the simulation of the water content is easier than in the simulation of the pwp distribution.

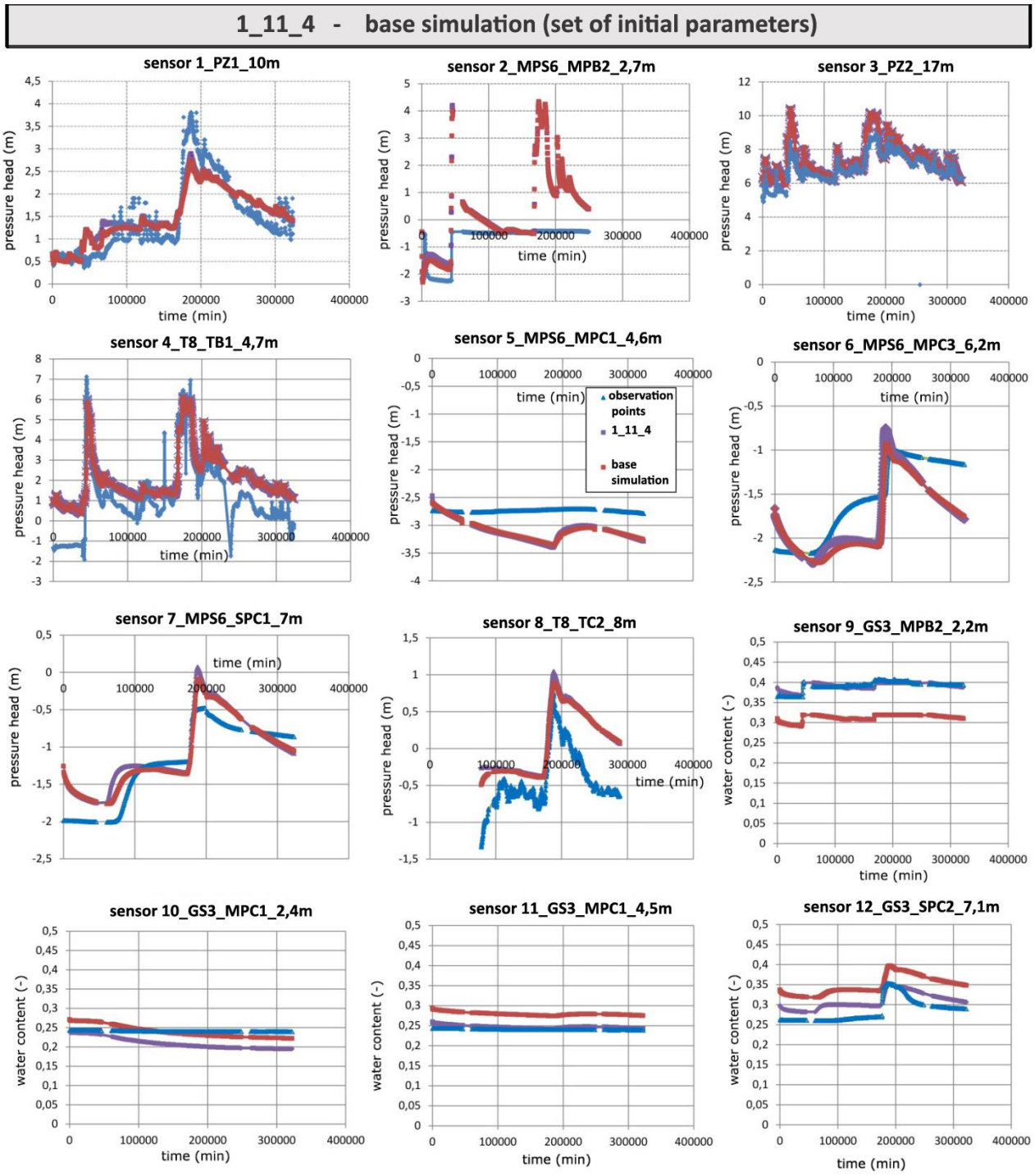


Figure 29: Graphs presenting the comparison between the investigated indirect simulation (purple points) with the observed dataset (blue points) and the base simulation (red points) for each sensor. The indirect simulation 1_11_4 use a dataset of observation points of type ''wc''.

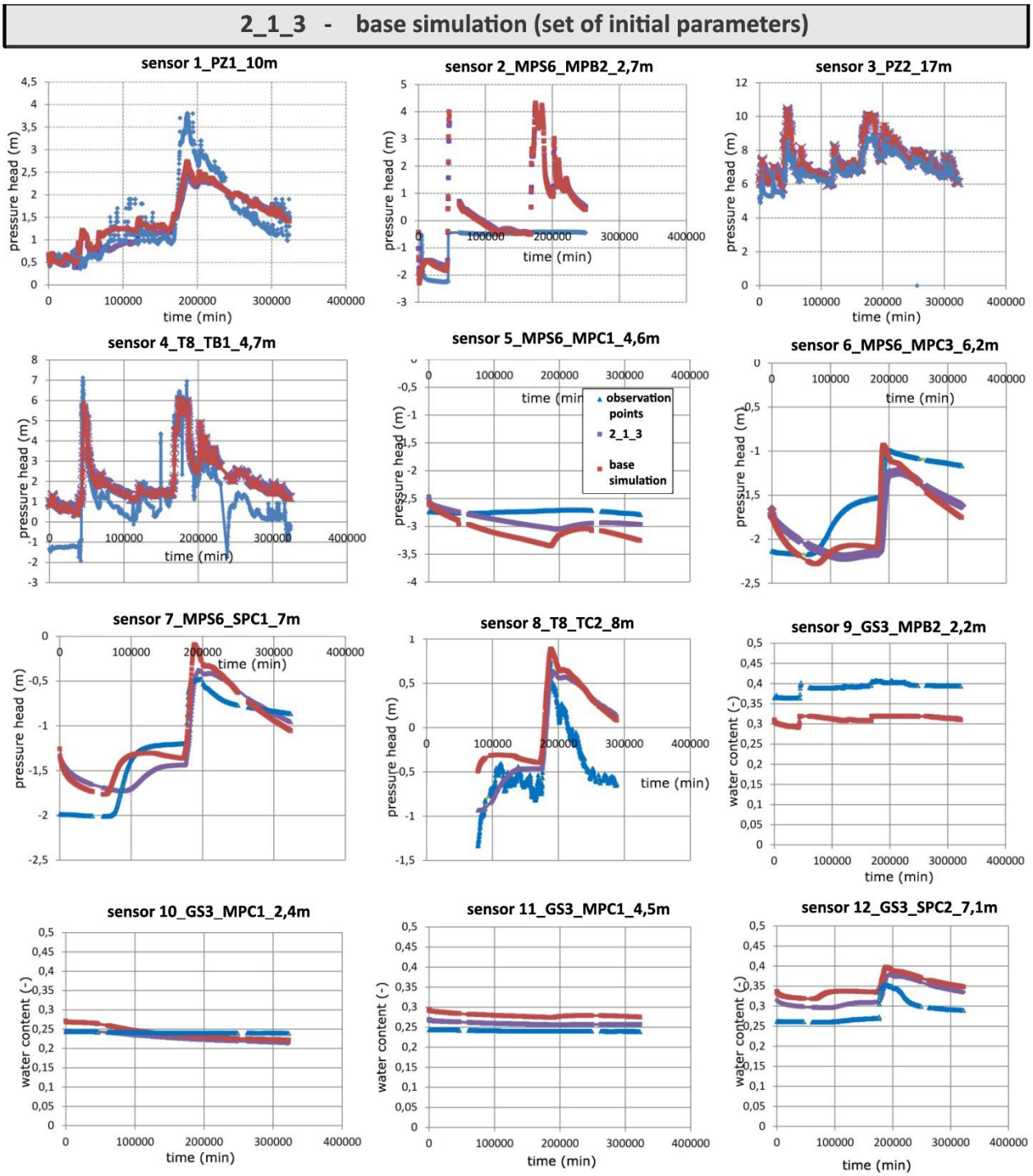


Figure 30 Graphs presenting the comparison between the investigated indirect simulation (purple points) with the observed dataset (blue points) and the base simulation (red points) for each sensor. The indirect simulation 1_11_4 use a dataset of observation points of type ''PH''.

Despite a statistical analysis of the optimized parameters obtained after the 1st and 2nd calibration phases is not legitimate due to a non-uniformity of the conditions in which optimized parameters have been obtained (as explained extensively at the beginning of chapter §7_5), a standard normal distribution could be used to visualize the dispersion of the optimized values (blue dots) and of the parameters obtained by laboratory tests (red dots) for each parameter of the hydraulic models VGM and HVGM.
This representation could give interesting information on the dispersion of the optimized parameters with respect to:

- the type of observed datasets used in the inverse analysis (PH, wc, PH+wc),

- the initial guess (set of initial parameters)

The VGM+2cm model gives an overall problem of convergence in the indirect problem and a poor performance of the available inverse simulations which resulted in a small number of accepted simulations after the 1st and 2nd calibration phases. In fact the magnitude of the residuals in the objective function depends largely on the proper selection of the hydraulic model and the use of an inappropriate model results in an increase of the deviation between simulated-observed data. The VGM+2cm, as observed previously for the HVGM+2cm, is not suited for the presented case study and problems of convergence, non-uniqueness of the solution are the main consequences that are clearly observed in the indirect problem. For this reason, the VGM and HVGM are the only hydraulic models to be investigated and to whom sets of optimized parameters will be proposed.

The normal distribution has been presented in for the VGM, and in for the HVGM. For the VGM, the curve is not presented for the *n* parameter belonging to the fluvial and foundation layers because no optimized values are available (no simulations have overcome the second calibration phase). For the HVGM the curve is not presented for *ϑr* and *n* (fluvial layer) and *n* (foundation layer). For the aquifer layer, the saturated permeability Ks is the only optimized parameter for obvious reasons.

A red line is used for the mean of the optimized values while the green lines represent the confidence interval. A black dashed line represents the initial value of the parameter.

If we examine closely , we can verify that the outermost values of the parameters (α, θs, Ks, θs in the different layers) on the Gaussian curve have been obtained by water content dataset, while the values closer to the peak of the Gaussian curve are all obtained by a pressure head dataset or a pressure head+wc dataset.

If we consider the simulations that use a PH+wc dataset, the optimized values are closer to the ones obtained with a PH dataset. This means that the inverse simulation is driven by the pressure head dataset and less by the water content dataset when both the datasets are considered together. A possible explanation could be that sensors monitoring water potential in the riverbank under investigation showed a greater change (larger interval of change) during the simulated period compared to sensors monitoring water content (as could be observed in and). The responsive sensors have a more relevant/effective role in parameter calibration.

Some examples of optimized parameters values obtained using different datasets of observation points are provided below.

-HVGM, embankment layer:
α=1,830 (simulation 1_2_12) with a wc dataset
α=1,609 (simulation 1_4_6) with a wc dataset
α=0,854 (simulation 1_15_3) with a PH+wc dataset
α=0,8537 (simulation 1_1_9) with a PH dataset
As could be observed, the parameter α optimized using a PH or a wc+PH dataset has almost the same value which is very similar to the initial value (0,850).

-VGM, embankment layer:
Ks=2,08*10^{-5} m/min (simulation 1_6_10) with a wc dataset
Ks =9,00*10^{-5} m/min (simulation 1_5_9) with a PH dataset
Ks =9,07*10^{-5} m/min (simulation 1_16_1) with a PH+wc dataset
As could be observed, the parameter Ks optimized using a PH or a wc+PH dataset has almost the same value which is very similar to the initial value (9,00*10^{-5} m/min).

The reason why a PH dataset leads to optimized parameters close to the initial values could stay in the high performance of the model when simulating a pwp distribution using the initial set of parameters and the relate difficulty to obtain an optimization of the same parameters using inverse analysis. This leads to optimized parameters values in close proximity of the initial ones.

In the first calibration phase, it has been observed a general poorer performance of the simulations that use a water content dataset with respect to simulations that use a PH dataset. For example for the embankment layer, the θs parameter optimized using a PH dataset (1_3_4) gives a R^2 value of 0,95 (1 best fit); a NSE value of 0,94 (1 best fit); a IA' of 0,89 (1 best fit), while the same parameter optimized with a water content dataset (1_4_3) gives a R^2 value of 0,72; a NSE value of 0,12; a IA' of 0,65 for the whole 2018 simulation period. Similar examples could be made looking at the inverse simulations of the remaining parameters. Despite this, as observed in , 9/11 indirect simulations that will be tested in the validation phase use a water content dataset. As a result the greatest enhancements are shown by the sensors monitoring the wc (see) and to a lesser extent by sensors monitoring PH because the simulation of the water content distribution has wider margins of improvement with respect to the distribution of PH. This should not lead to conclude that a dataset of type w.c. has to be preferred over a dataset PH or PH+wc. When available the use of all the three datasets (wc, PH and mixed) is recommended.

In the presented case study:

- the hydraulic parameters of the fluvial layer have been optimized using only a water content dataset (because only sensors monitoring wc are available within the layer) or a mixed dataset (PH+wc) when the fluvial layer is optimized with other layers (e.g. embankment+fluvial or embankment+foundation+fluvial)
- the foundation layer has been optimized using only a PH dataset (because only sensors monitoring water potential are available within the layer) or a mixed dataset when the foundation layer is optimized with other layers.
- the embankment layer has been optimized with both the datasets because both the typology of sensors (wc and water potential measurements) are available within the layer.

The quality of the final solution of the parameter estimation problem is strongly dependent on the observed dataset used in the objective function. The different types of dataset, the laboratory or in situ procedure with which the dataset has been obtained, the number of independently measured data, the identification of the type of dataset which is best suited for the estimation of the hydraulic conductivity function or of the soil water retention parameters (or both), are very relevant aspects to consider in order to obtain a reliable parameters inverse estimation (Hopmans et al, 2002).

Another interesting aspect that could be observed looking closely at the simulations that overcome the second calibration phase and optimize two hydraulic parameters at a time, is that the parameters increase/decrease with respect to the initial estimate depending on the input dataset. For example for the VGM, the parameters Ks and θs increase with respect to the initial estimate using a dataset of type PH+wc, they both decrease using a dataset of type wc, while θs increases and Ks decreases with a dataset PH. This behaviour underlines once again the influence of the input dataset in the inverse problem and the impossibility to identify a clear trend of increase/decrease of the hydraulic parameters with respect to the initial guess.

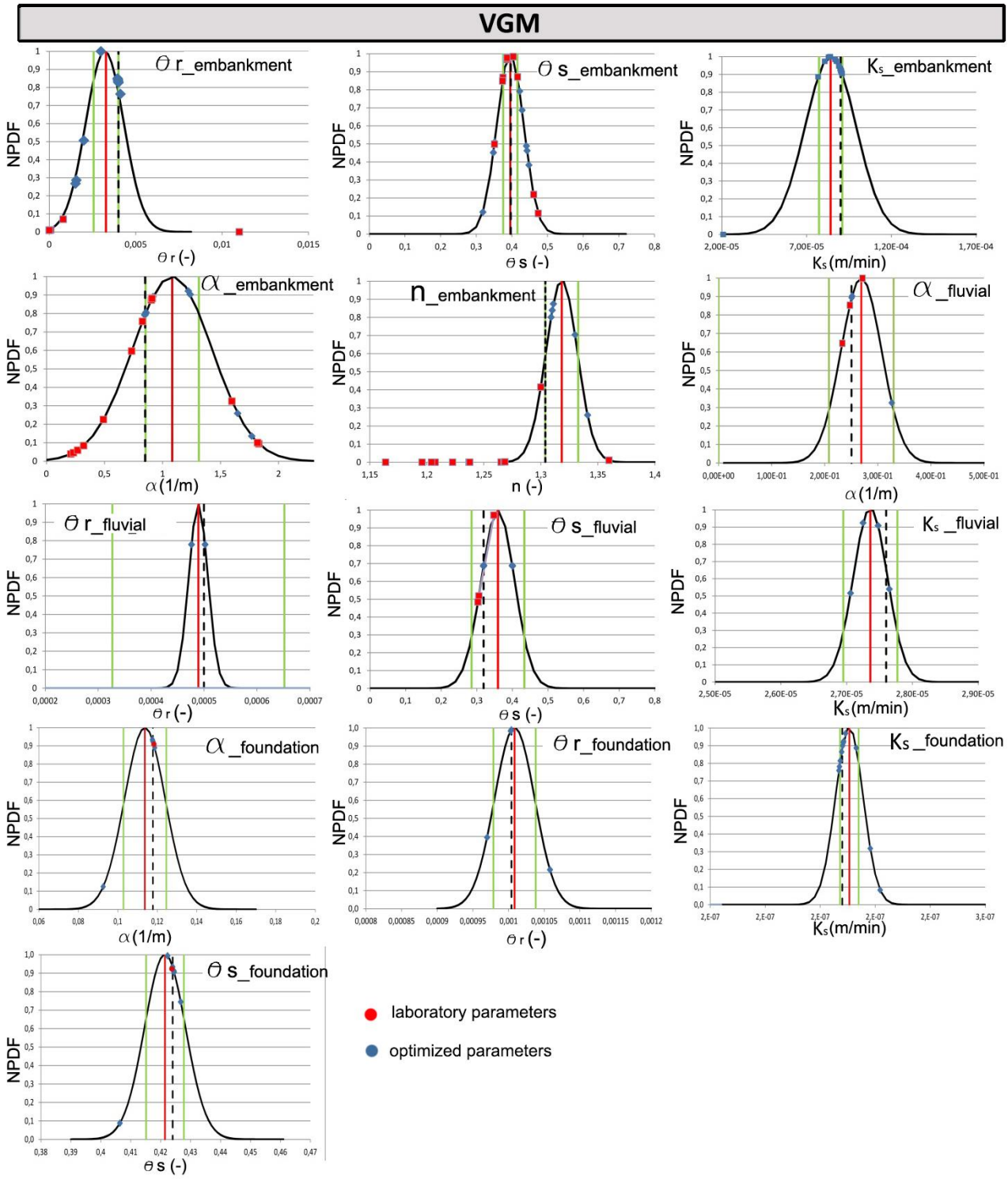


Figure 31: Normalized probability density function (NPDF) of the optimized parameters divided in layers (VGM). The hydraulic parameters obtained in laboratory tests (evaporation tests) are superimposed to the Gaussian curves (red dots).

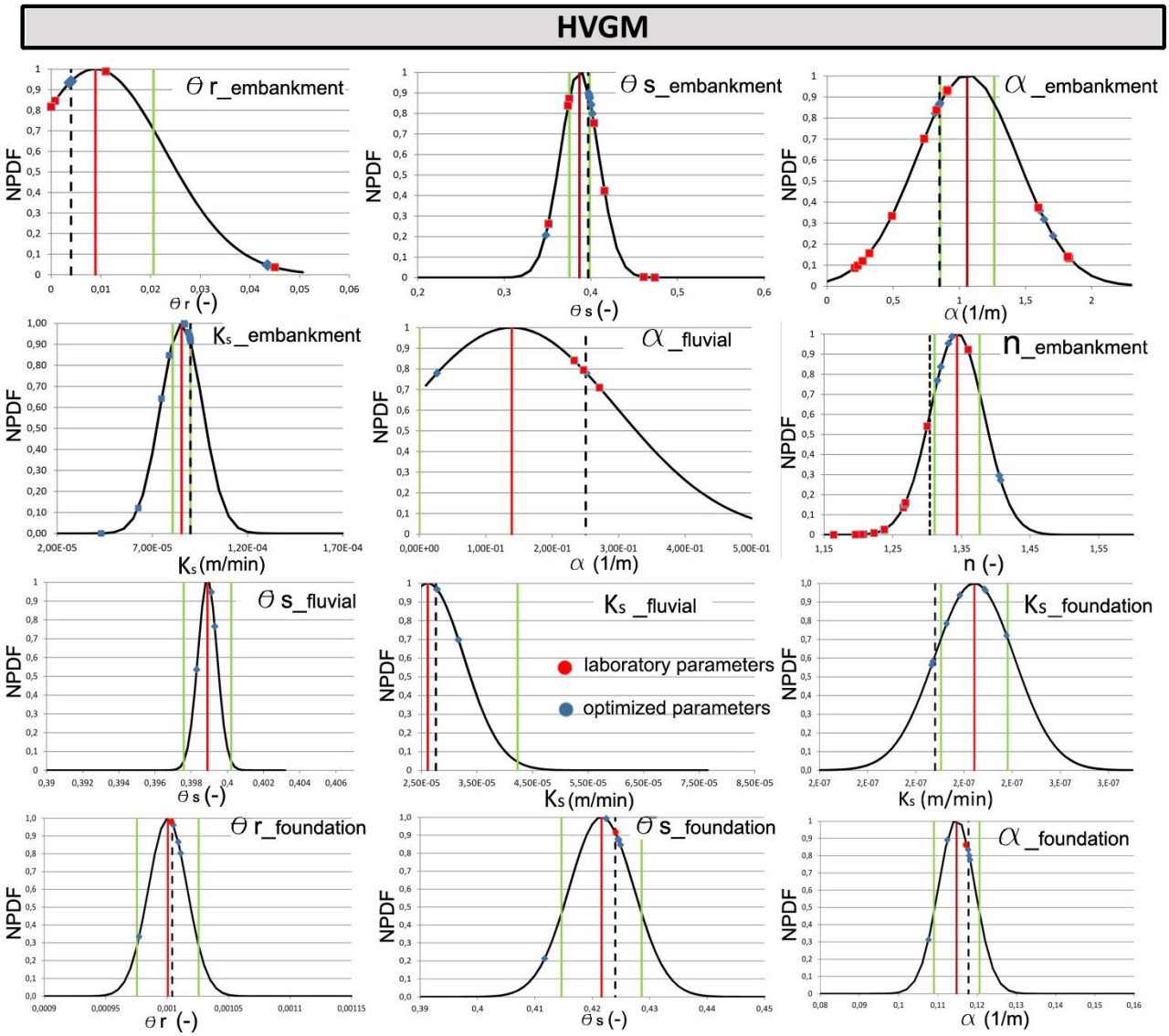


Figure 32: Normalized probability density function (NPDF) of the optimized parameters divided in layers (HVGM). The hydraulic parameters obtained in laboratory tests (evaporation tests) are superimposed to the Gaussian curves (red dots).

7_5_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The third calibration phase looks at the performance of the single sensors (observation points), while in the first phase the performance of the whole set of available sensors was under scrutiny. N° 12 sensors have been chosen for this purpose. N° 81 indirect simulations overcome the second calibration phase and they will be tested in the third phase. As in the first calibration phase the discrepancy between observed-simulated datasets is investigated and the same set of metrics/indices is used. The performance is investigated on the whole simulation period and on the two main flood events of the 2018 simulation period. The final goal is the individuation of the best possible match between in situ data and simulated data in the observation points. To do so metrics/indices are compared to the values of the base simulation which uses the initial set of hydraulic parameters in order to quantify the obtained enhancement in the model performance. Greater the difference between optimized simulation and base simulation, greater the enhancements obtained.

The considered twelve sensors have different weights in the indirect problem because located in different positions in the bank body and because they measure different physical properties (wc /PH). The sensors located in the saturated zone have a minor weight compared to the ones in the unsaturated one. At the same way the sensors monitoring wc have a minor weight compared to sensors monitoring PH, due to the fact that the final purpose of the calibration is the obtaining of a reliable pwp distribution in the most relevant time instants. Moreover different weights are attributed to the set of adopted metrics/indices in order to have a balanced score between accuracy and efficiency metrics. It has been observed that three sensors over 12 do not show improvements in any considered indirect simulation. The indirect simulations that perform best are 9: 5 simulations for the HVGM and 4 simulations for the VGM. Moreover the whole group of nine indirect simulations optimizes the parameters of the embankment layer and embankment+fluvial layers using a wc dataset and a weighting distribution equal to unity for the whole dataset. The simulation of the wc distribution has wider margins of improvement with respect to the pwp distribution because the initial set of parameters assures a good initial simulation of the PH observed dataset. Moreover it has been observed that the majority of the indirect simulations enhances the performance of some sensors and reduces the performance of other observation points; no simulation is able to improve the performance of all the considered sensors together. For this reason for each problem it is important to individuate which sensors have a greater relevance in order to understand on which sensors to concentrate the calibration procedure. For the presented case study, four sensors installed in the embankment layer and one positioned in the berm are the most significant to elaborate a reliable pwp map for future stability analysis. It is important to point out that the enhancements in the performance of the model in proximity of the considered sensors (observation points) are of small entity due to the fact that the initial set of parameters comes from reliable laboratory tests which assure a good performance of the base simulation. This helps to save time in the calibration phase avoiding to perform a high number of indirect simulations in multi-layered models. The optimized parameters are located in a narrow interval close to the initial value (from laboratory). The achievement of the highest enhancement has been obtained for sensors monitoring the water content with respect to sensors monitoring the water potential because the base simulation shows a very good performance in the simulation of the pwp distribution and a poor performance in the simulation of the wc distribution. The enhancement of a model which has a very good initial performance is more difficult. At the beginning of the third calibration phase the dispersion of the optimized parameters which have overcome the first two phases have been investigated. A normal distribution has been used and the mutual position of optimized parameters, initial parameter, confidence intervals has been investigated on the same curve. It has been observed that the optimized parameters close to the peak of the curve (where the initial parameter is located) have been obtained with a pressure head dataset or a mixed dataset (PH+wc) while the optimized parameters obtained using a water content dataset are on the external sides of the curve. This means that the optimized parameters obtained with a PH dataset differ only slightly from the initial guess which is consequence of the good performance of the base model (using the initial set of parameters). Moreover the optimized parameters obtained using a mixed dataset (wc+PH) are closer to the parameters obtained using only a PH dataset and different from the parameters obtained using a wc dataset. This means that inverse analysis is driven more by a PH dataset and this could be due to the larger variance/changes in the pwp trend observed in the 2018 simulation period (with respect to the wc trend). Responsive sensors have a more relevant/effective role in parameter calibration. In addition it has been observed that optimized parameters increase/decrease with respect to the initial guess depending on the observed dataset in use; this underlines once again the relevance of the typology of observed dataset on the obtained results.

A proper statistical analysis of the obtained optimized parameters is not legitimate because indirect simulations have not been run under the same conditions (same hydraulic model, same weighting distribution, same dataset of observation points).

In the following chapter 7_6, different initial sets of parameters have been tested in order to understand their influence on the results of the inverse analysis. The calibration campaign already exposed has used a unique set of initial parameters, the average values of the hydraulic parameters obtained by laboratory tests.

7_6 INVERSE ANALYSIS STARTING FROM A DIFFERENT SET OF INITIAL PARAMETERS

As already reported, the inverse analysis performed for the presented case study started from the same set of initial parameters, given by the average values of the hydraulic parameters obtained by laboratory tests. This choice has been made in order to have the same starting point for all of them. This allows to compare under the same conditions the performed indirect simulations.
We can thus focus our attention on:

- the role of the observed dataset (different type of dataset),
- the combination of hydraulic optimized parameters
- the different weighting distributions for the observed dataset.

It is well known that the Levenberg-Marquardt non linear parameter estimation is strongly dependent on the initial parameter estimate which represents the starting point of the first iteration of the process.
For this reason, it has been hypothesized that laboratory tests were not available to estimate the initial set of hydraulic parameters. Following Schaap et al (2001), the pedotransfer functions incorporated in the software package ROSETTA in Hydrus 2D have been used to predict the initial values of the hydraulic parameters by means of a neural network model. The Rosetta pedotransfer function estimates the unsaturated hydraulic properties based on surrogate soil data such as the grain size distribution of the material or the textural classes, the bulk density and/or the water content correspondent to 33 kPa and 1500 kPa.

At this point, it is interesting to investigate the behaviour of the objective function starting from a different point of the parameter space (different initial guess of the hydraulic parameters). We want to detect if the minimum of the objective function is a global minimum or a local minimum and the influence of the initial estimate of the parameters on the results of the indirect simulation.

In order to reduce the number of indirect simulations to perform, the embankment is the only layer whose hydraulic parameters are optimized by means of inverse anaysis and the initial values of hydraulic parameters have been computed by Rosetta software. No other changes in the initial conditions, boundary conditions, mesh size and mesh refinements have been introduced in order to allow an easier comparison with the previous simulations.
The input observed dataset to perform the inverse analysis is composed of 8 sensors installed in the embankment layer:

- 4 sensors monitoring the water potential (T8_TC2_8m, MPS6_MPC1_4,6m, MPS6_SPC1_7m, MPS6_MPC3_6,2m)

- 4 monitoring the water content (GS3_SPC2_7,1m, GS3_MPC1_4,5m, GS3_MPC1_2,4m, GS3_MPC3_6,4m).

The hydraulic model chosen is the van Genuchten Model in order to have the minimum number of parameters to optimize.
The convergence of the model under the new set of initial parameters for the embankment layer (while all the other layers have remained unchanged) has been extremely difficult.
The calculation of the direct problem using the new set of parameters for the embankment layer proved to be extremely slow and difficult and a very small number of inverse simulations have reached the convergence. No conclusions could be drawn from such a small number of indirect simulations.

For this reason a second attempt has been made: the parameters that are going to be optimized using inverse analysis have been changed while the rest of them remains as the initial guess (from laboratory tests). For example to optimize the α parameter of the embankment layer, its value is changed from 0,85 to 0,5 1/m while the rest of the parameters remains unchanged. The new set of initial parameters are reported in the fourth row in Table 48.
51 indirect simulations have been performed, 17 using a pressure head dataset, 17 using a water content dataset and 17 using a water content+pressure head dataset. Combination of parameters, from a single parameter to all the hydraulic parameters together, have been optimized as done in the previous calibration programme.
13 indirect simulations have succeded. Then the optimized set of hydraulic parameters have been tested in the direct problem using the whole set of observation points (sensors monitoring wc + sensors monitoring water potential). This was performed similarly in the previous calibration programme in order to compare the performance of the successful indirect simulations under the same conditions. Changing only the initial value of the parameter that is going to be optimized allows to understand the direction in which the objective function is moving (increasing /decreasing/remain the same) with respect to the initial guess of the parameter and it is possible to compare this value with the one obtained by laboratory tests.

		embankment					
		teta r (-)	teta s (-)	alfa (1/m)	n (-)	ks (m/min)	l (-)
simul. number	initial parameters (from laboratory tests)	0,004	0,397	0,850	1,304	9,00E-05	0,5
	new initial parameters	0,100	0,200	0,500	1,100	9,00E-06	0,5
1	α_PH	0,0040	0,397	0,501	1,304	9,00E-05	0,500
2	α_wc	0,0040	0,397	1,674	1,304	9,00E-05	0,500
3	α+k_PH	0,0040	0,397	0,504	1,304	9,08E-06	0,500
4	α+θs_PH	0,0040	0,201	0,499	1,304	9,00E-05	0,500
5	k_PH	0,0040	0,397	0,850	1,304	9,02E-06	0,500
6	k_wc	0,0040	0,397	0,850	1,304	4,10E-05	0,500
7	k_wc+PH	0,0040	0,397	0,850	1,304	4,86E-10	0,500
8	θr_PH	0,1004	0,397	0,850	1,304	9,00E-05	0,500
9	θs_PH	0,0040	0,201	0,850	1,304	9,00E-05	0,500
10	θs_wc	0,0040	0,346	0,850	1,304	9,00E-05	0,500
11	θs+θr_PH	0,1004	0,201	0,850	1,304	9,00E-05	0,500
12	θs+k_wc+PH	0,0040	0,348	0,850	1,304	4,03E-07	0,500
13	θs+θr_wc+PH	0,0124	0,425	0,850	1,304	9,00E-05	0,500

Table 48: The table reports the initial set of hydraulic parameters from laboratory tests (third row), the new set of initial parameters (fourth row) and the optimized set of hydraulic parameters of the 13 successful indirect simulations performed. In red the changed values with respect to the initial guess.

The sets of optimized parameters of the 13 successful indirect simulations are reported in Table 48: in red the optimized parameters and in black the unchanged initial parameters.

Looking closely at the Table 48, it is possible to observe that:

- the optimization of the parameter ''n'' has been insuccessfull.
- θr in simulations 8 and 11 remains very close to the initial guess (0,1), while in simulation 13, θr has decreased of an order of magnitude getting closer to the value obtained using laboratory tests (0,004).
- θs in simulations 11, 9, 4 remains very close to the initial guess (0,2), while in simulations 10, 12 it increases to values closer to the laboratory one (0,397) and in simulation 13 it increases above the laboratory value. There is a clear trend of increase toward greater α values for all the successful simulations.
- In simulations 1, 3, 4, α remains close to the initial guess (0,500) while in simulation 2 it reaches a value extremely high (1,674) (higher than the one obtained in laboratory tests).
- the saturated permeability Ks remains very close to the initial guess ($9*10^{-6}$ m/min) in simulations 3 and 5; while in simulations 12 and 7 it decreases to values far below the one hypothesized initially ($9*10^{-5}$ m/min). It is possible to state that there is not a clear trend of increase/decrease of Ks with respect to the initial guess.

From all this, we can state that the initial parameters dataset has a huge influence on the results of the optimization process. In few cases as for θs (simulation 10) and Ks (simulation 6), the optimized parameters are very close to the values obtained by laboratory tests while in the majority of the cases, as α+θs (simulation 4) and θs+θr (simulation 11), the optimized parameters stay in close proximity to the intial values, without apporting a relevant change in the performance of the observation points. Moreover we can observe that the majority of the successful inverse simulations have been carried out using a pressure head dataset but the optimized parameters obtained differ only slightly from the initial guess while the greatest changes in terms of value of the parameters (compared to the initial guess) are shown by the simulations that use a wc dataset or a wc+PH dataset. The same was observed in the previous chapter §7_5 for the other programme of inverse analysis performed.
All the calibration phases have been performed on the 13 indirect simulations in order to evaluate in a quantitative way their performance.

The calibration phases show numerically that the performances of all the thirteen simulations are poorer compared to the base simulation (that uses the initial parmeters obtained by laboratory tests) with the ecception of the simulation 2 that optimizes the α parameter (wc dataset). The score attributed to each indirect simulation in the first, second and third calibration phases are reported in Table 49.

The indirect simulation 2 is the only one that shows an enhancement of the performance with respect to the base simulation. Figure 33 shows the performance in the observation points/monitoring points located in the embankment using the set of optimized parameters from simulation 2: the simulated dataset is reported in blu, the base simulation (parameters from laboratory tests) in green and the observed dataset in red. As could be observed in Figure 33 the simulated performance of the sensors monitoring water content has been optimized while the simulated performance of 3 out of 4 sensors monitoring the pressure head is poorer (or the same) compared to the simulation that uses the set of initial parameters from laboratory tests (MPS6-MPC1-4,6m is the only sensor measuring water potential whose simulated performance is enhanced). The indirect simulation 2 will be tested in the validation phase together with the inverse analysis selected in chapter §7_5 because it has overcome all the three calibration phases.
As observed also in chapter §7_5, it has proven to be more difficult to optimize the performance of sensors of water potential compared to sensors monitoring water content, and this could be related to the reduced changes in their value registered during the simulation period.

sim. number	NEW SET OF INDIRECT SIMULATIONS	1° CALIBRATION PHASE	2° CALIBRATION PHASE	3° CALIBRATION PHASE
1	α_PH	5,54	4,00	-6946,50
2	α_wc	**7,67**	**4,00**	**2401,68**
3	α+k_PH	5,75	5,00	-8874,49
4	α+θs_PH	5,74	3,75	-31532,87
5	k_PH	5,75	3,25	-1302,95
6	k_wc	3,60	1,50	119,47
7	k_wc+PH	5,31	3,00	-2732,10
8	θr_PH	5,74	3,50	-5836,21
9	θs_PH	5,73	3,00	-33131,61
10	θs_wc	3,92	4,25	442,62
11	θs+θr_PH	5,73	4,75	-15503,14
12	θs+k_wc+PH	5,29	5,00	2391,17
13	θs+θr_wc+PH	5,29	4,00	-8343,67
MAX SCORE		15	8	-
MINIMUM SCORE		7,5	4	> 0

Table 49: Scores attributed to the first, second and third calibration phases of the new set of indirect simulations. Only the indirect simulation 2 shows a good performance in all the three phases.

We can conclude that the majority of the simulations remain in their initial local minima while few of them approach the minimum in which the base simulation (initial parameters from laboratory tests) is located. It is incorrect to state that the ''laboratory'' parameters (base simulation) are located in the global minimum of the parameter space, what it is possible to say is that the simulations analysed in chapter §7_5 (third phase of the calibration process) show a better performance compared to the 13 simulations analysed in the present chapter and for this reason the new thirteen combinations of optimized parameters are located for sure in a local minimum.

It is interesting at this point to observe where the optimized values of the present chapter §7_6 are located in the gaussian distribution of the optimized parameters obtained from the previous calibration phase (inverse simulations using the set of ''laboratory'' initial parameters). As could be observed α and θs are arranged on both sides of the gaussian curve showing to be close to the optimized parameters obtained starting from the ''laboratory'' initial parameters. For what concerns Ks and θr, they are located outside the normal distrubution and far from the parameters obtained previously from the calibration procedure. This is a further evidence of the incidence of the chosen set of initial parameters on the optimization procedure. For this reason it is always suggested when dealing with gradient type minimization technique to repeat the minimization using different initial estimate of the parameters in order to be sure that the final solution is located in a global minimum and not in a local minimum (Simunek and Hopmans, 2002).

It has been observed a general difficulty in the convergence of the inverse analysis starting from a set of initial parameters far from the ''good'' one. While working on the mesh, iteration criteria and time step control, it is possible to obtain the convergence of the direct problem using a set of first attempt initial parameters, the same is not assured for the indirect simulations. A ''smooth'' convergence of the direct simulation is necessary but not sufficient for a ''smooth''convergence of the inverse analysis performed on the model. This problem is related to the complexity of the case study that requires always an update of the model when the initial set of hydraulic parameters is changed. For this reason with each new iteration of the objective function, the set of parameters ''change'' more and more from the initial one and the convergence becomes increasingly difficult (long calculation time due to the necessity of a high number of

iterations at each time instant of the simulation). When the inverse analysis starts from a ‘‘good’’set of intial parameters, few iterations are required to adjust the set to the optimal one and no convergence problems are detected: as a result the optimized set of parameters differs only slightly from the initial one.
The presented case study falls within this framework since the performed inverse analysis starts from a good set of reliable hydraulic parameters that come from SWRC obtained in multiple laboratory experiments. The convergence of the direct simulation has been obtained quite easily after having observed in this phase the relevance of the minimum time step[3] on the model convergence. The convergence of the indirect simulations has been obtained quite easily and the high number of successful simulations allowed to perform and in-depth calibration of the parameters. The individuation of the conditions (dataset of observation points and weighting distributions to use, set of parameters to be optimized together) for which a good performance of the inverse analysis is guaranteed, allows to reduce the computational efforts of the calibration phase (high number of indirect simulations to be performed) when a complex hydraulic-geotechnical problem is adressed.

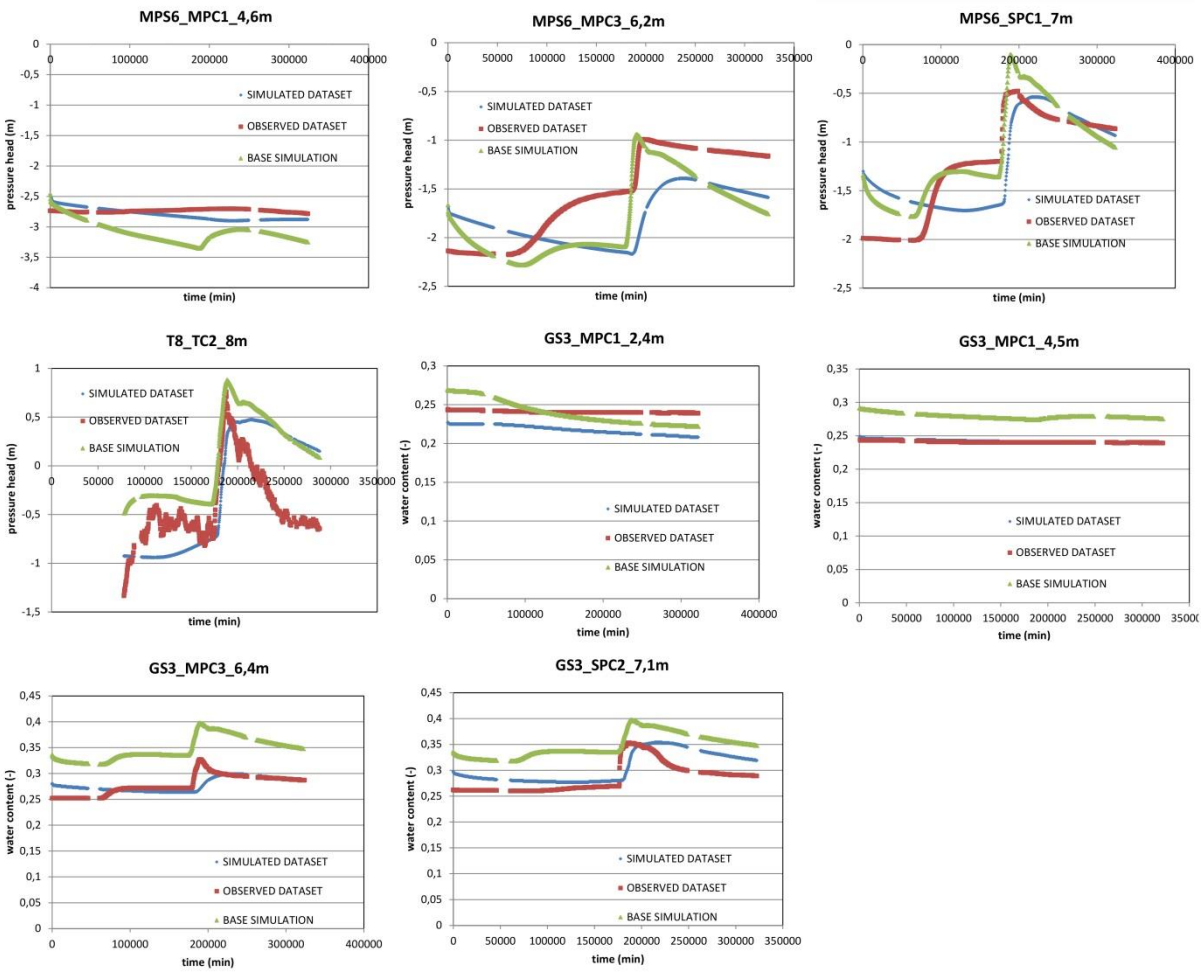


Figure 33: Graphs showing the simulated dataset of the indirect simulation 2 (blue dots) together with the base simulation that uses the set of initial parameters from laboratory tests (green dots) and the observed dataset (red dots).

[3] Minimum permitted value of time increment (for more information see chapter 6)

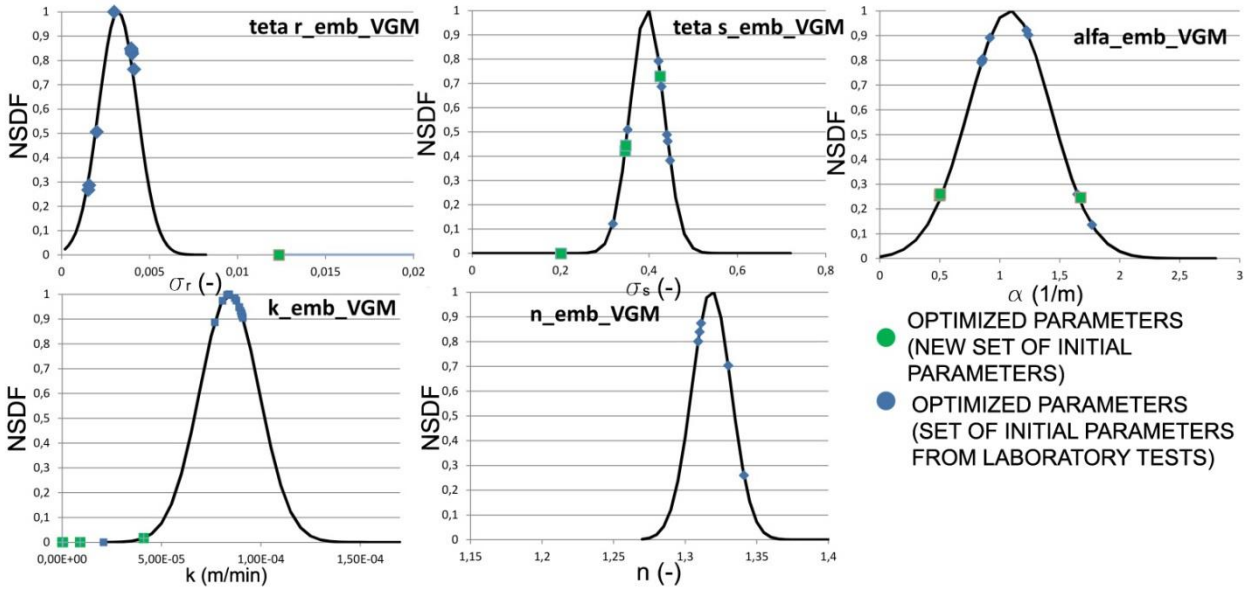


Figure 34: Graphs showing the Gaussian distribution of the optimized parameters using the new sets of initial parameters (chapter 7_6) and the optimized parameters obtained in the previous chapters using a set of initial parameters from laboratory tests are superimposed on the same curves.

7_6_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The Levenberg-Marquardt algorithm is strongly dependent on the initial estimate of the model parameters which influence the starting point of the iterative process in the indirect problem. Chapter 7_6 focuses on the investigation of the influence of the initial set of parameters on the results of the inverse analysis. In other words the behaviour of the objective function starting from different points in the parameters space has been investigated in order to understand if the set of optimized parameters obtained are located in a local minimum or in a global one. At the same way, it is interesting to observe in which direction the objective function is moving with respect to the initial guess of the parameters (incremental/ decremental /same local minimum). In the previous chapters which report the calibration procedure in three phases, a unique set of initial parameters has been used: the average values of the hydraulic parameters from laboratory tests. In chapter 7_6 pedotransfer functions have been used to predict hydraulic parameters starting from physical properties of the investigated soil. To reduce the number of indirect simulations to perform, the unique set of hydraulic parameters that has been changed is the one belonging to the embankment layer. The VGM is chosen as unique hydraulic retention model to adopt because it requires the lower number of parameters to optimize. The dataset of observation points used in the inverse problem is composed of 8 sensors: 4 monitoring the pwp and 4 monitoring the wc in the embankment layer. It has been observed that the convergence of the indirect simulations is quite difficult using the set of initial parameters obtained by pedo-transfer functions. The reduced number of converged simulations makes impossible to draw any conclusions. For this reason, it has been tried to change only the initial values of the parameters that are going to be optimized in each simulation, from one to all the hydraulic parameters of the embankment layer, in order to ease the indirect simulation convergence.13 simulations have reached convergence. The majority of the successful simulations use a PH dataset but they show small differences with respect to the initial guess of the parameters. Major differences are observed in correspondence of a wc dataset of observation points. The farer the initial set of parameters is from the ''optimal one'', the greater the difficulties in the convergence of the model. The use of an initial set of hydraulic parameters from laboratory tests makes fast and easy the convergence both of the direct and indirect problems and it reduces the necessity of an in-depth calibration phase of the parameters. In fact,

as observed in the considered case study the obtained enhancements after and in-depth and long calibration phase are reduced, especially for the simulation of the pwp distribution. It is also relevant to report that an easy convergence of the direct problem is necessary but not sufficient to obtain an easy convergence of the indirect problem.

The successful sets of optimized parameters have been plotted on a Gaussian curve in order to investigate how they have changed with respect to the initial guess of each hydraulic parameter. The α parameter shows an increase with respect to the initial guess, any simulation optimizing n has been successful, Ks has no clear trend of increase/decrease with respect to the initial guess, θs and θr show one part of the optimized values close to the initial guess from pedo-transfer functions and others close to the initial guess from laboratory tests. We can conclude that the majority of the simulations stay in the local minima of the initial value, others approach the values of the parameters obtained in laboratory tests..

The quantification of the indirect simulations performance by means of the calibration procedure in three phases has been applied to the thirteen successful simulations. One single indirect simulation has a better performance with respect to the base simulation with the initial set of parameters (from laboratory tests). This simulation uses a wc dataset and it will be tested in the validation phase together with the other indirect simulations selected in the previous chapters (from 7_2 to 7_5).

In chapter 7_7 the role of the weighting distribution on the results of the indirect simulations is investigated. Different typologies of weighting distributions have been tested and the performance of the indirect simulations carried out investigated by means of the calibration procedure in three phases presented in the previous chapters.

7_7 WEIGHTING DISTRIBUTIONS FOR THE DATASET OF OBSERVATION POINTS IN THE INDIRECT PROBLEM

The role of the weighting distributions attributed to the observed dataset used in the objective function of the indirect problem has been extensively discussed in chapter §2_7 of the present thesis. In the calibration programme performed in the previous chapters, two weighting distributions have been chosen for the observed dataset: equal to 1 or according to a TMA distribution.

This chapter mainly focuses on the use of different weighting distributions for the observed dataset based on:

- the different temporal instants in which the observed data are recorded (n°6 different weighting distributions)
- the type of sensor, accuracy of sensors and location in the bank body (n°3 different weighting distributions).

The details of the different cases addressed are presented extensively in Table 50. Each group of inverse analysis (identified with the code 2_XX with XX varying from 1 to 9) is composed of 13 indirect simulations with a variable number of optimized hydraulic parameters (from 1 to 5). The hydraulic parameters to be optimized belong only to the embankment layer and the chosen hydraulic model is the VGM because it requires a low number of parameters. The groups of inverse analysis from 2_1 to 2_6 use a PH dataset while groups from 2_7 to 2_9 a dataset of type PH+wc.

Table 50 presents in the third column the details of the sensors used in the inverse problem and in the second column details on the weighting distributions adopted for each group of inverse analysis. The groups of inverse analysis 2_1 and 2_2 use weighting distributions that reduce the relevance of the off-peak periods with respect to the flood peak periods.

For example the group 2_1 uses a weight=0 for the off-peak periods and the group 2_2 a weight one-half of the one attributed to the peak periods.

The groups of inverse analysis 2_3; 2_4 and 2_5 use weighting distributions which scale the weights based on the water level and flood persistence according to the procedure exposed in the first calibration phase in which partition coefficients for each flood and off-flood periods were computed depending on the area underneath the curve in the graph time – river water level.

In particular the group of inverse analysis 2_3 uses a weight=0 for the off-peak periods while the peak periods are scaled according to duration - water level, the group 2_4 scales the weights of both the periods (peak/off-peak) and the maximum weight for both the periods is 1; the group 2_5 scales the weights of both the periods (peak/off-peak), the maximum weight for the peak periods is 1 and the maximum weight for the off-peak periods is 0,5.

The group of inverse analysis 2_6 gives more weight to the observation points during the flood drawdown (max weight drawdown=1, max weight up to the flood peak=0,5) because most of the times, the maximum variation of the phreatic surface and the minimum stability conditions (minimum FOS) occur with some delay with respect to the peak of the flood (maximum water level reached).

For the groups of inverse analysis 2_7; 2_8 and 2_9, weighting distributions based on the different locations, type and accuracy of sensors are adopted. Group 2_7 gives more weight to the sensors monitoring water potential with respect to the sensors monitoring wc because the final aim of the modelling is the elaboration of a pwp distribution in the more relevant time laps to be used for future reliable stability analyses. In group 2_8, the weights are assigned based on the position of the sensors in the section: the weights of the sensors in the unsaturated zone are set equal to 1, the weights of the sensors in the saturated zone are set equal to 0,5 and the weights of the shallow sensors (depth < 3m from the soil surface) are set equal to 0,5. The sensors located in the unsaturated zone of the bank section are the most relevant for the type of analysis that is being conducted in order to see the effect of the transient river water level on the degree of saturation and soil matric suction. A reduction of the weight is applied to the sensors located in the unsaturated zone but with a distance <3m from the soil surface because these sensors are influenced predominantly by root water uptake (plants transpiration contribution) and evaporation from the soil surface and both the phenomena are dealt with approximations due to the lack of experimental data. The group of inverse analysis 2_9 gives a major weight to the sensors with greater accuracy (as T8 with respect to MPS6) and to sensors monitoring water potential (with respect to sensors monitoring wc).

The percentage of successful indirect simulations that uses a weighting distribution based on the different temporal periods is equal to 37% (with respect to the initial number) while 33% for the simulations that use location, type and accuracy of the sensors. These percentages are much lower compared to the weighting distribution equal to 1 for all the observed data (60%) and almost the same percentages obtained for the TMAT (30%) (see Table 23).

The groups of inverse analysis were analysed using the same calibration procedure in three phases presented in the previous chapters in order to investigate the quality of the simulations output. This could give interesting insights on the use of different weighting distributions. The procedure of evaluation in three phases will not be reported in this chapter because no innovative aspects are introduced (for more information see chapters §7_3, §7_4 and §7_5 of the present thesis).

In Table 51 the results in terms of scores attributed to the first, second and third calibration phases are reported for each group of inverse analysis. In red the two indirect simulations that overcome the calibration phases and that will be tested in the validation phase on a new simulation period. Both the simulations (2_1_3 and 2_3_3) optimize the α parameter of the embankment layer using weighting distributions that consider equal to zero the weights of the off-peak periods.

Table 51 shows that all the considered indirect simulations overcome the first calibration phase which is less ''selective'' with respect to the following ones, as seen in the previous chapters when applied to the other set of indirect simulations. Six indirect simulations overcome the second calibration phase and in Table 52 the scores attributed to each sensor on the whole simulation period are reported. As could be observed relevant enhancements in the performance of the sensors are registered only for 2_1_3 and 2_3_3 for the sensors MPS6_MPC1_4,6m (PH) T8_TC2_8m (PH) and GS3_MPC1_4,5m. In fact it is important to remember that a total score in the third phase between 0-5 underlines enhancements impossible to detect in the graph observed dataset, base simulation and optimized simulation vs time.

WEIGHTING DISTRIBUTIONS BASED ON THE DIFFERENT TEMPORAL PERIODS		
2_1	WEIGHT OF THE DATASET DURING THE FLOOD PEAK PERIODS=1 WEIGHTS OF THE DATASET DURING THE OFF-PEAK PERIODS=0	PH dataset: Optimization of the hydraulic parameters of the embankment layer
2_2	WEIGHT OF THE DATASET DURING THE FLOOD PEAK PERIODS=1 WEIGHT OF THE DATASET DURING THE OFF-PEAK PERIODS=0,5	
2_3	WEIGHT OF THE DATASET SCALED ON THE BASIS OF DURATION AND HYDROMETRIC LEVEL REACHED IN THE FLOOD PEAK PERIODS (max weight=1) WEIGHT OF THE DATASET IN THE OFF-PEAK PERIODS=0	
2_4	WEIGHT OF THE DATASET SCALED ON THE BASIS OF DURATION AND HYDROMETRIC LEVEL REACHED IN THE FLOOD PEAK AND OFF-PEAK PERIODS (max weight=1 for both the periods)	
2_5	WEIGHT OF THE DATASET SCALED ON THE BASIS OF DURATION AND HYDROMETRIC LEVEL REACHED IN THE FLOOD PEAK AND OFF-PEAK PERIODS (max weight=1 for the peak-periods and 0,5 for the off-peak periods)	
2_6	WEIGHT OF THE DATASET IN THE FLOOD PEAK PERIODS SCALED ON THE BASIS OF DURATION AND HYDROMETRIC LEVEL REACHED (max weight=0,5 for the dataset up to the peak and max weight=1 for the drawdown weight of the off-peak periods=0; weight off-peak periods=0)	
WEIGHTING DISTRIBUTIONS BASED ON THE DIFFERENT LOCATION, TYPE AND ACCURACY OF SENSORS		
2_7	WEIGHT OF THE SENSORS MONITORING WATER POTENTIAL=1; WEIGHTS OF THE SENSORS MONITORING WATER CONTENT=0,5	PH+wc dataset: Optimization of the hydraulic parameters of the embankment layer
2_8	WEIGHT BASED ON THE POSITION OF THE SENSORS IN THE SECTION: WEIGHT OF THE SENSORS IN THE UNSATURATED ZONE=1; WEIGHTS OF THE SENSORS IN THE SATURATED ZONE=0,5, SHALLOW SENSORS (< 3m)=0,5	
2_9	WEIGHT BASED ON THE ACCURACY OF THE SENSORS. WEIGHT ASSOCIATED TO T8 TENSIOMETERS=1; TO MPS6 SENSORS=0,75; TO GS3 AND SM150T=0,6	

Table 50: The table is a summary of the different groups of inverse analysis that have been performed using different weighting distributions based on the different time periods or on the different location.

	OPTIMIZED PARAMETERS EMBANKMENT	1st PHASE	2nd PHASE	3rd PHASE WHOLE SIMULATION PERIOD	3rd PHASE FLOODS "C" and "G"
2_1_2	k	11,1	2,50	9,5	-21,7
2_1_3	α	11,5	4,00	602,4	1026,3
2_1_4	θs	11,1	3,25	-1508,8	-3421,5
2_1_5	n	11,8	3,00	40,3	137,1
2_1_7	k, θs	11,1	2,75	-899,1	-2143,4
2_1_12	θr	11,8	2,00	-0,2	1,1
2_2_2	k	11,4	3,00	9,4	-21,7
2_2_5	n	12,1	3,00	34,4	112,1
2_2_7	k, θs	11,4	2,25	-648,9	-1592,4
2_2_12	θr	12,1	1,50	15,3	44,5
2_3_2	k	11,3	2,75	11,7	-27,3
2_3_3	α	11,5	4,10	590,9	1000,8
2_3_4	θs	11,4	2,25	-1428,5	-3265,7
2_3_5	n	12,1	2,25	33,8	112,7
2_3_13	θr,k	12,1	2,25	17,7	55,3
2_4_2	k	11,3	3,25	11,7	-27,3
2_4_5	n	11,4	3,50	167,1	1306,1
2_4_12	θr	12,1	2,50	7,0	15,8
2_5_2	k	11,4	3,25	14,5	-38,8
2_5_3	α	11,6	4,50	-482,7	1315,2
2_5_4	θs	11,4	3,75	-3985,0	-3985,0
2_5_5	n	11,4	2,50	519,2	519,2
2_5_9	α,θs	12,1	3,50	-1262,9	-1262,9
2_5_12	θr	12,1	2,25	55,9	55,9
2_7_3	α	13,7	3,50	-527,6	734,4
2_7_4	θs	13,6	3,25	-2271,7	-2975,0
2_7_5	n	13,5	3,00	-934,9	152,3
2_7_12	θr	13,5	1,50	-981,6	2,2
2_9_1	k	12,0	3,00	-952,6	-113,2
2_9_3	α	13,2	3,50	-301,9	1328,4
2_9_5	n	12,0	3,50	-1108,1	1303,4
2_9_6	k,α	14,4	4,00	-1067,5	662,9
2_9_7	k, θs	12,2	4,00	-3282,4	-5036,2
2_9_12	θr	12,3	2,00	-963,5	55,3
2_9_13	θr, k	12,0	3,00	-956,0	-135,2
2_8_2	k	12,0	3,75	-949,1	-133,8
2_8_9	α, θs	13,7	4,50	-5858,9	-5722,2
2_6_2	k	11,1	2,50	-969,5	-29,0
2_6_3	α	11,3	3,25	-395,9	988,8
2_6_4	θs	11,1	3,50	-2279,9	-2996,6
2_6_5	n	11,6	3,00	-928,0	175,9
2_6_7	k, θs	11,1	2,75	-2365,9	-3162,6
Threshold score		7,5	4	>0	>0

Table 51: Scores attributed to each indirect simulation in the first, second and third calibration phases. The second column reports indication of the parameter/s of the embankment layer optimized in each indirect simulations.

The indirect simulations 2_1_3 and 2_3_3 show enhancements in the performance of the sensors both in the whole simulation period and during the floods ''C'' and ''G''. The groups of inverse analysis 2_6, 2_7; 2_8 and 2_9 show the lowest accuracy with no enhancements registered in the third calibration phase (whole period) for any indirect simulation. It is interesting at this point to represent the optimized parameters obtained using the different weighting distributions (blue points) on a Gaussian curve together with the hydraulic parameters from laboratory SWRCs (orange points). In red the initial parameter (initial guess of the inverse analysis). In Figure 35 the normalized probability density functions of the hydraulic parameters of the embankment layer are presented. As could be observed:

- θr parameter: The majority of the points (laboratory and optimized parameters) are located on the left side of the curve, indicating a reduction in the parameter value with respect to the initial guess. The dispersion of both the values (laboratory and optimized parameters) is small as the confidence interval.
- θs parameter: The majority of the laboratory parameters and optimized parameters from inverse analysis are located on the apex of the Gaussian curve that indicates a small dispersion of the parameters and, for the same reason, a small confidence interval.
- n parameter: the optimized parameters and the laboratory parameters are located on the opposite side of the Gaussian curve with respect to the initial guess. Laboratory parameters suggest a reduction in the value of ''n'' with respect to the initial guess while the optimized parameters an increase in the parameter value. The laboratory parameters present a wide dispersion of their values. The optimized parameter n=1,43 on the right side of the curve belongs to the group of inverse analysis 2_9 (weighting distribution based on type of sensor, position and accuracy) that showed together with 2_7 and 2_8 a poor performance.
- α parameter: the optimized parameters and the laboratory parameters are located on the opposite side of the Gaussian curve with respect to the initial parameter. The optimized parameters suggest an increase in the parameter value while the laboratory parameters a decrease with respect to the initial guess. Few laboratory parameters are located on the apex of the Gaussian curve. Wide dispersion of the data for optimized and laboratory parameters. The optimized parameters α=2,54 and α=2,66 on the right side of the curve belong to the groups of inverse analysis 2_8 and 2_9 (weighting distributions based on type of sensor, position and accuracy) that showed together with 2_7 poor performances.

SCORE (WHOLE SIMULATION PERIOD) $\Delta_{i\ total}*w_i$													
	PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 4,6m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 SPC2 7,1m	TOT SCORE
2_1_3	-0,05	-0,09	0,03	-0,14	182,99	-0,28	-0,27	13,74	0,43	-8,86	412,98	1,88	602
2_3_3	-0,04	-0,05	-0,07	-0,15	178,88	-0,28	-0,25	14,23	0,41	-5,99	402,36	1,84	591
2_5_3	0,33	-9,00	-0,41	-0,96	-107,75	-0,61	0,44	-59,35	-25,60	-69,82	-208,29	-1,63	-483
2_9_6	0,25	-9,47	-0,48	-1,19	-11,87	-0,24	0,17	-65,00	-24,88	-719,07	-236,14	0,44	-1067
2_9_7	0,30	-8,74	-0,27	-0,94	-250,17	-1,14	0,26	-44,22	-25,89	-635,22	-2307,39	-9,02	-3282
2_8_9	0,27	-9,64	-0,34	-1,09	-11,16	-0,01	0,31	-76,77	-25,07	-3324,39	-2406,77	-4,20	-5859

Table 52: The table presents the values of $\Delta_{i\_tot}*w_i$ (for more information see chapter §7_5) of each indirect simulation that overcome the second calibration phase. In pink the values >0 that indicate an enhancement in the performance of the sensor.

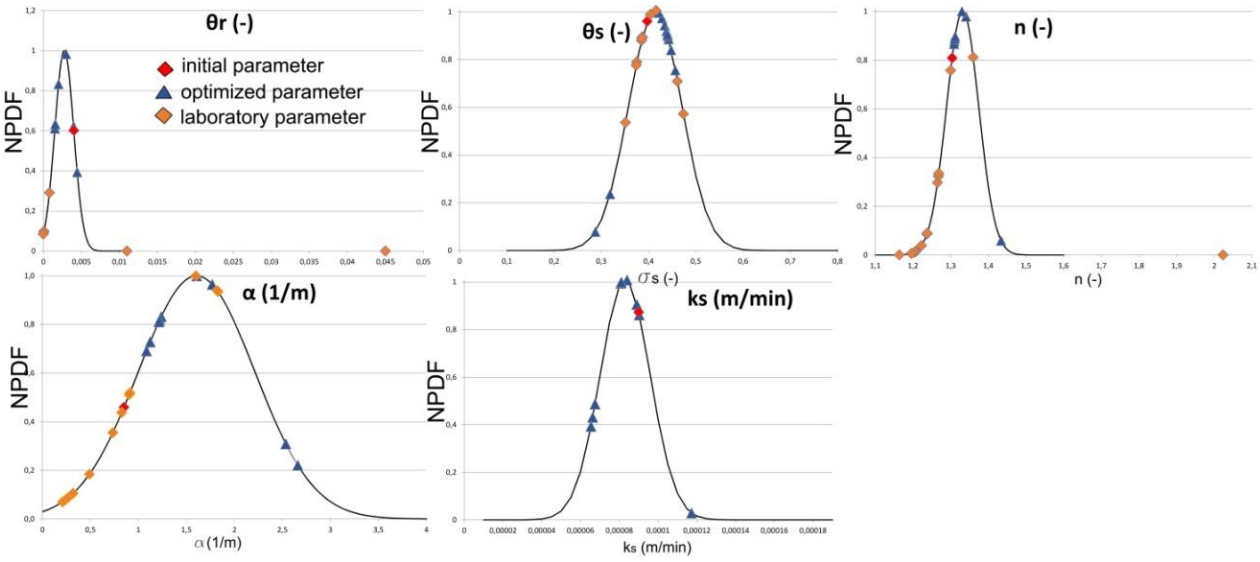


Figure 35: Gaussian distributions of the optimized parameters (blue dots) that use different weighting distributions with indication of the parameters from laboratory SWRC (orange dots) and of the initial guess (red dot).

7_7_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

In the calibration phase carried out in the previous chapters, two different weighting distributions have been tested: all the observed data with the same weight equal to unity and weights set according to the Triangular Moving Average Technique (TMAT). In chapter 7_7 different weighting distributions have been tested in order to investigate their influence on the results of the indirect problem. The chosen weighting distributions are based on the different temporal instants in which the observed data are recorded (flood peaks /off-peaks/drawdown), based on the type of sensors, the accuracy of the instruments and the location of each of them in the bank body. For each weighting distribution, 13 indirect simulations with a variable number of optimized parameters belonging to the embankment layer have been carried out. The VGM has been chosen as the unique calibrated hydraulic model because it requires the lower number of parameters. Two different observed datasets have been used, a PH dataset and a mixed dataset (PH+wc). The weighting distributions based on the temporal periods are the following: a weight of zero assigned to the data belonging to the off-peak periods; a weight equal to half (0,5) of the max weight given to the flood periods (1);a weight scaled based on river water height and duration of each flood (weight off-peak period equal to 0); the weights of both flood and off-flood periods scaled based on river water level and duration, the same of the previous case but with a max weight for the flood periods of 1 and for the off-flood periods of 0,5.;weights of 1 for the data belonging to the drawdowns and 0,5 for the data up to the flood peak. As could be easily understood the flood periods result more interesting for the purposes of the modelling we are performing because the most critical stability conditions for riverbanks are registered during flood events and in particular for fine grained riverbanks during flood events with medium-high river water levels and more persistent in time. In some cases it has been observed that the minor FOS are registered after the peak of the flood event, during rapid drawdowns. The weighting distributions based on the sensors are the followings: major weight is given to the sensors monitoring PH (1) with respect to sensors monitoring wc (0,5); major weight is given to sensors located in the unsaturated zone (1) with respect to the ones located in the saturated zone (0,5) an exception is done for the sensors located in the unsaturated zone with a depth <3m from the soil surface (greater influence of the evapo-transpiration phenomena) to which a weight of 0,5 is set; major weight is given to sensors with greater accuracy (for example the weight of sensor T8 is 1 and of sensor MPS6 is 0,75). It has been observed that 37% of the simulations using weighting distributions based on the different temporal periods have reached convergence while 33% of the simulations using weighting distributions based on type, accuracy and position of the sensors. In the previous calibration programme, 60% of the simulations using a weighting distribution equal to unity have reached convergence, 30% using a TMAT. The reason behind this could be easily understood: the difficulty of the indirect simulations to reach convergence increases if a major weight is given to the observed data during the most significant time periods (flood peaks) while the optimization of the hydraulic parameters result easier in the remaining periods (off-flood) when no particular changes are detected by the installed sensors. From all this we clearly observe the relationship between adopted weighting distribution and results of the inverse analysis. The methodological approach in three phases already presented in the previous chapters has been adopted to quantify objectively the performance of the successful simulations. Only two indirect simulations have overcome the three phases. Both the indirect simulations optimize the α parameter of the embankment layer using two weighting distributions based on the different temporal periods. Looking at the dispersion of the optimized parameters on a Gaussian distribution with respect to the initial guess and with respect to the parameters from laboratory tests, it has been observed that the optimized θr show a reduction with respect to the initial guess (average of laboratory parameters); optimized θs are located in the confidence interval of the laboratory parameters (peak of the Gaussian

curve); while for n and α, the optimized parameters and the laboratory parameters are located on the opposite side of the Gaussian curve with respect to the initial guess.

In the next chapter 8, the indirect simulations that have overcome the three calibration phases are tested in the validation phase on a new monitoring period, the 2019. The validation phase applies the same methodology seen for the calibration phase to quantify the performance of the model simulations using the optimized set of parameters already identified in the calibration phase.

8 MODEL VALIDATION

The model validation is the process used to determine to what extent the model corresponds to the real system. Validation is an inductive process at the end of which the modeller could draw conclusions on the accuracy of the model prediction systematically comparing the model output to independent observations of the real phenomenon. A simple and widely used procedure to perform a validation process is the ''split-sample'' procedure. This procedure is used to perform the validation of hydrological models due to the uncertainties affecting these models (Arsenault et al, 2018). This procedure consists of two groups of real data: one group used for the calibration process and the second one used to emulate a real world phenomenon (''out of sample'' application). Sometimes the groups are inverted and calibration and validation are repeated on switched datasets (Simunek et al, 2012). A model is considered valid for a set of experimental dataset if its accuracy is within an acceptable range. The accuracy required has to be set prior to the development of the model according to the model intended purpose. The majority of the time the validation or not of the model depends on a subjective decision based on personal evaluation during model creation and based on the results of various tests. A more objective way to evaluate the model performance is using scores. Scores are determined subjectively on various aspects of the validation process, then the scores are summed and the model is considered valid if its overall and category scores result greater than a threshold.

For the presented case study, the 2017-2018 period (from 12th November 2017 to the 23rd June 2018) was used to calibrate the model parameters while the 2018-2019 period between 24th June 2018 and 30th June 2019 is used to validate the set of optimized parameters on a new monitoring period (''first seen'' data). The 2018-2019 period is composed of 4 months of low water level (up to the 27th October 2018) and 7 months of high water level (4 flood events from 27th October 2018 to 24th February 2019, 5 flood events from 23rd March 2019 to the end of May 2019) in order to test the model in both the different conditions. It is relevant to report that the investigated period is characterized by five events of medium relevance (flood A; flood B; flood F; flood G; flood H), one event of low relevance (flood C) and three events of high intensity (flood D; flood H; flood L). Floods F,G,H,I,L could be read as a single flood (flood M) composed of different peaks-drawdowns of the river water level. In Figure 36 the hydrometric water level of the 2018-2019 period is presented and in Table 53 the duration of the single flood and off-flood events are reported with their start and end time (in minutes and days). To model the validation period, no changes have been done to boundary conditions and geometry with respect to the 2018 period (calibration period); as initial conditions the pressure head distribution of the last time lapse of the calibration period has been imported.

The sensors considered in the validation phase are: PZ2_10m, MPS6_MPB2_2,7m, PZ1_17m, T8_TB1_4,7m, MPS6_MPC1_3,05 m, MPS6_MPC1_4,6m; T8_TC1_4,7m; MPS6_MPC3_6,2m; MPS6_SPC1_7,0m;

T8_TC2_8m; GS3_MPB2_2,2m; GS3_MPC1_2,4m; GS3_MPC1_4,5m; GS3_MPC3_6,4m and GS3_SPC2_7,1m. No information are available on the hydrometric level in Ponte Motta section in the period 24/2/19 to 23/3/19 due to data acquisition problems. For this reason, for the numerical modelling in Hydrus2D, a linear distribution of hydrometric level has been hypothesized in the gap (point line in Figure 36). Moreover GS3_MPB2_2,2m; GS3_MPC1_2,4m; GS3_MPC1_4,5m and GS3_SPC3_7,1m do not have data in the interval 24/3/19 - 18/6/19 due to a logging issue.

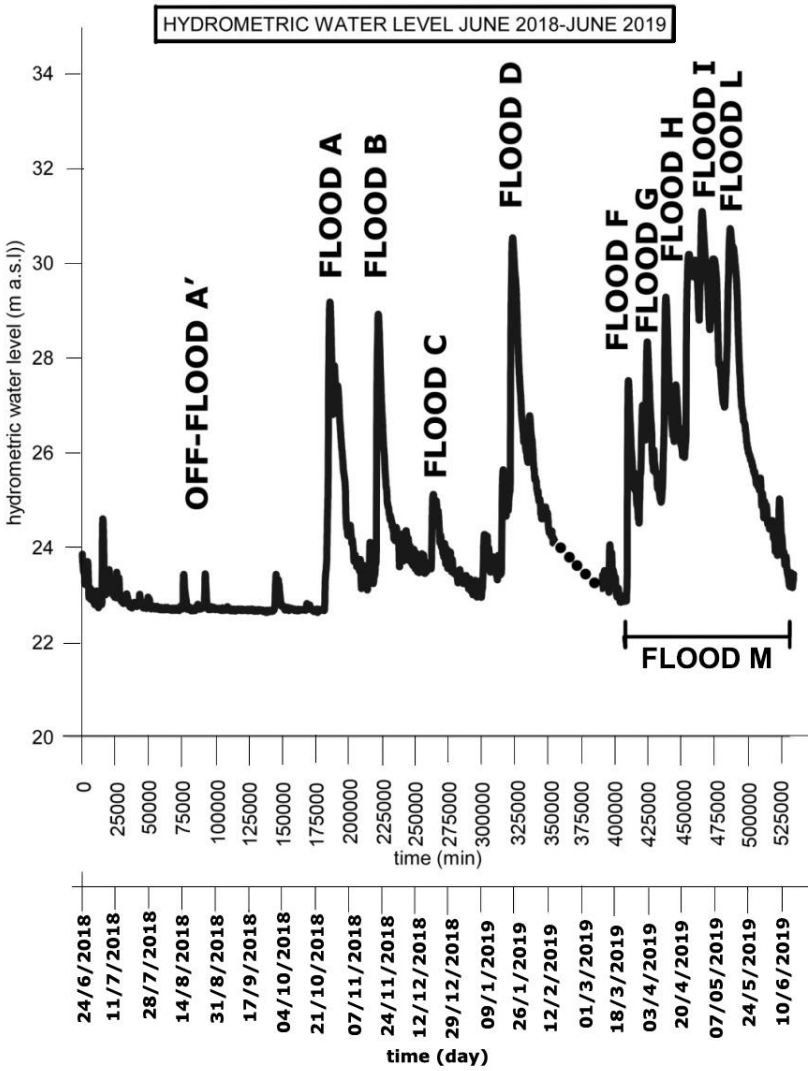


Figure 36: Hydrometric water level of the temporal period from June 2018 to June 2019. On the x-axis the scales in minutes (used in the Hydrus software for the modelling) and in days are reported. Floods are named with letters from A to L (flood periods) and A' for the initial period of low river water level (24/6/2018 to 28/10/2018).

	2019_VALIDATION PERIOD			
	FROM (min)	TO (min)	FROM (day)	TO (day)
FLOOD A	182280	213630	28/10/2018	19/11/2018
FLOOD B	214500	249870	19/11/2018	14/12/2018
FLOOD C	258870	278910	20/12/2018	03/01/2019
FLOOD D	311910	352980	26/01/2019	24/02/2019

FLOOD F	408330	417390	29/04/2019	05/05/2019
FLOOD G	417390	433920	05/05/2019	17/05/2019
FLOOD H	433920	450630	17/05/2019	28/05/2019
FLOOD I	450630	481230	28/05/2019	18/06/2019
FLOOD L	481230	544190	18/06/2019	18/06/2019
FLOOD M (F+G+H+I+L)	408330	544190	29/04/2019	18/06/2019
OFF-FLOOD A'	0	182280	24/06/2018	28/10/2018

Table 53: The table reports the initial and final time instants (in days and minutes) of each considered flood event.

The procedure in three phases that will be used in the validation phase to evaluate the performance of indirect simulations is the same explained and applied in the previous chapters, for this reason it will not be reported here again (for more information see chapters §7_3, §7_4 and §7_5). The performances of 10 indirect simulations are tested in the validation phase: 8 indirect simulations that overcome the third calibration phase in chapter §7_5, 1 indirect simulation that overcome the third calibration phase in chapter §7_6 (indirect simulations starting from another set of initial parameters) and 2 in chapter §7_7 (indirect simulations starting from different weighting distributions). A list of the chosen indirect simulations is reported for sake of clarity in Table 54:

INDIRECT SIMULATION IDENTITY CODE	HYDRAULIC MODEL	LAYER	TYPE OF OBSERVED DATASET	OPTIMIZED PARAMETERS	WEIGHTING DISTRIBUTION	CHAPTER
1_2_2	HVGM	embankment	w.c.	α	weights=1	8_5
1_2_6	HVGM	embankment	w.c.	Ks, α	weights=1	8_5
1_2_12	HVGM	embankment	w.c.	Ks, α, θr	weights=1	8_5
1_4_6	HVGM	embankment	w.c.	Ks, α	weights=1	8_5
1_4_7	HVGM	embankment	w.c.	Ks, θs	weights=1	8_5
1_11_3	VGM	emb+fluvial	w.c.	α	weights=1	8_5
1_11_4	VGM	emb+fluvial	w.c.	θs	weights=1	8_5
2_1_3	VGM	embankment	PH	α	Flood peak periods=1; Off-peak periods=0	8_7
2_3_3	VGM	embankment	PH	α	Flood peak periods= weights dependent on water height and persistence; Off peak-periods=0	8_7
Simulation 2	VGM	embankment	w.c.	α	weights=1	8_6

Table 54: The table reports the 10 indirect simulations that will be tested in the validation phase with indication of the hydraulic model, layer, type of observed dataset, weighting distribution used to obtain the set of optimized parameters for each of them. The last column reports the chapter where more information on each indirect simulation could be found.

Before showing the results of the performance evaluation in three phases of the considered indirect simulations, an investigation of their behaviour with respect to the base simulations (VGM or HVGM using the set of initial parameters) is provided. In Table 55 the results of the application of accuracy and

efficiency metrics to the flood peak periods D and M (the most relevant of the validation period), to the off-peak period (A') and to the whole simulation period (June 2018-June 2019) is presented. In pink the values of the metrics that show an enhancement with respect to the base simulation. As it is possible to observe:

- the flood periods D and M are well represented by the indirect simulations and the set of metrics/indices shows to be close to the best fit (see last row in Table 55). Simulation 1_11_4 is an exception. The values of the efficiency metrics (IA, IA_j, NSE_j, NSE, IA', $IA_{j=1}$) are closer to the best fit with respect to the accuracy metrics (MAE, RMSE) indicating a good-high performance of the investigated simulations. We understand once again the importance of considering a set of metrics of different typologies which are able to highlight different aspects of the datasets under investigation.
- The off-flood period A' (the more relevant for temporal duration) is poorly represented by the chosen indirect simulations (compare the values of the metrics with the best fit) especially by simulations 1_11_3 and 1_11_4.
- Globally the performance of the indirect simulations over the whole period is medium-poor and it results halfway between the performance during the flood peak periods and the off-peak periods. Even in this case, simulations 1_11_3 and 1_11_4 show performances poorer than the ones of the base simulations. It is important to stress the relevance of a good representation of the flooding events for the purpose of the topic we are dealing with, as seen in Table 55 (flood D and M).

	FLOOD D									FLOOD M								
	R^2	MAE	RMSE	IAmod	IA	IAJ	NSEj	NSE	IA'	R^2	MAE	RMSE	IAmod	IA	IAJ	NSEj	NSE	IA'
1_2_2	0,892	0,893	0,616	0,831	0,968	0,831	0,619	0,856	0,810	0,834	0,834	0,872	0,780	0,952	0,780	0,514	0,789	0,757
1_2_6	0,893	0,894	0,605	0,833	0,968	0,833	0,626	0,859	0,813	0,832	0,832	0,882	0,777	0,952	0,777	0,508	0,786	0,754
1_2_12	0,890	0,892	0,608	0,833	0,967	0,833	0,624	0,855	0,812	0,833	0,833	0,869	0,780	0,952	0,780	0,515	0,789	0,757
1_4_6	0,894	0,896	0,592	0,836	0,969	0,836	0,634	0,863	0,817	0,830	0,830	0,892	0,775	0,951	0,775	0,502	0,784	0,751
1_4_7	0,895	0,898	0,675	0,819	0,965	0,819	0,583	0,836	0,792	0,825	0,824	1,012	0,751	0,945	0,751	0,435	0,747	0,718
1_11_3	0,873	0,876	0,644	0,824	0,961	0,824	0,602	0,827	0,801	0,827	0,826	0,941	0,766	0,948	0,766	0,475	0,768	0,737
1_11_4	0,891	0,894	0,802	0,794	0,955	0,794	0,504	0,774	0,752	0,824	0,823	1,038	0,749	0,940	0,749	0,421	0,712	0,710
2_1_3	0,894	0,896	0,661	0,822	0,966	0,822	0,592	0,843	0,796	0,834	0,834	0,917	0,771	0,950	0,771	0,488	0,773	0,744
2_3_3	0,893	0,896	0,659	0,822	0,966	0,822	0,592	0,843	0,796	0,834	0,834	0,913	0,772	0,950	0,772	0,491	0,775	0,745
sim.2	0,892	0,893	0,617	0,831	0,968	0,831	0,619	0,855	0,809	0,835	0,835	0,875	0,779	0,952	0,779	0,512	0,788	0,756
Base VGM	0,892	0,895	0,778	0,798	0,957	0,798	0,519	0,790	0,760	0,825	0,824	1,029	0,750	0,941	0,750	0,426	0,720	0,713
Base HVGM	0,892	0,895	0,778	0,798	0,957	0,798	0,519	0,790	0,760	0,825	0,824	1,029	0,750	0,941	0,750	0,426	0,720	0,713
BEST FIT	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1
	OFF-FLOOD A'									WHOLE PERIOD								
	R^2	MAE	RMSE	IAmod	IA	IAJ	NSEj	NSE	IA'	R^2	MAE	RMSE	IAmod	IA	IAJ	NSEj	NSE	IA'
1_2_2	0,199	0,214	2,183	0,656	0,309	0,656	0,455	0,116	0,728	0,229	0,235	1,282	0,712	0,502	0,712	0,469	0,215	0,735
1_2_6	0,200	0,215	2,181	0,656	0,310	0,656	0,456	0,117	0,728	0,230	0,235	1,284	0,711	0,502	0,711	0,468	0,216	0,734
1_2_12	0,204	0,219	2,178	0,656	0,313	0,656	0,457	0,120	0,728	0,233	0,238	1,278	0,713	0,504	0,713	0,471	0,218	0,735
1_4_6	0,201	0,216	2,180	0,656	0,311	0,656	0,456	0,118	0,728	0,230	0,236	1,286	0,710	0,502	0,710	0,468	0,216	0,734
1_4_7	0,207	0,222	2,199	0,652	0,321	0,652	0,451	0,126	0,726	0,229	0,235	1,342	0,702	0,512	0,702	0,444	0,220	0,722
1_11_3	0,237	0,252	2,380	0,627	0,380	0,627	0,406	0,168	0,703	0,261	0,267	1,424	0,694	0,550	0,694	0,410	0,253	0,705
1_11_4	0,194	0,209	2,228	0,648	0,316	0,648	0,444	0,123	0,722	0,219	0,225	1,380	0,698	0,514	0,698	0,428	0,214	0,714
2_1_3	0,198	0,213	2,192	0,654	0,311	0,654	0,453	0,118	0,727	0,227	0,233	1,305	0,709	0,506	0,709	0,460	0,216	0,730
2_3_3	0,198	0,213	2,191	0,654	0,311	0,654	0,453	0,118	0,727	0,227	0,233	1,304	0,709	0,506	0,709	0,460	0,216	0,730
sim. 2	0,198	0,213	2,183	0,656	0,309	0,656	0,455	0,116	0,728	0,229	0,235	1,282	0,712	0,502	0,712	0,469	0,215	0,735
Base VGM	0,196	0,211	2,226	0,649	0,316	0,649	0,445	0,122	0,722	0,221	0,227	1,374	0,699	0,513	0,699	0,431	0,215	0,716
Base HVGM	0,196	0,211	2,226	0,649	0,316	0,649	0,445	0,122	0,722	0,221	0,227	1,374	0,699	0,513	0,699	0,431	0,215	0,716
BEST FIT	1	0	0	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1

Table 55: The table presents the values of the set of metrics (accuracy and efficiency) applied to the flood periods D and M, to the off-flood A' and to the whole simulation period. In pink the values of the metrics/indices that show an enhancement with respect to the values of the base simulation (VGM and HVGM in the last two row of the table)

To evaluate the performance of the indirect simulations, phases 1, 2, 3 are applied. To do so, the partition coefficients in Table 56, the different weights given to the sensors investigated in the analysis (of chapter §7_5) and the different weights given to the chosen metrics /indices (of chapter §7_5) are used. No differences are introduced with respect to the protocol applied in the previous chapters.

PARTITION COEFFICIENTS										
flood A	flood B	flood C	flood D	flood F	flood G	flood H	flood I	flood L	flood M (F+G+H+I+L)	SUM (A+B+C+D+M)
0,088	0,073	0,023	0,159	0,031	0,071	0,087	0,265	0,203	0,658	1

Table 56: Partition coefficients with which the score is subdivided in the different temporal periods (flood peak periods and off peak periods). For more information see chapter §7_3.

Table 57 reports the scores of the 1st, 2nd and 3rd phases of the performance evaluation for each indirect simulation. As it is possible to observe, all the indirect simulations overcome the first and second phases (scores higher than the imposed thresholds) while in the third phase 3/10 simulations show an insufficient performance (1_4_7; 1_11_3 and 1_11_4). These simulations do not overcome the validation procedure and they will be excluded from the final sets of optimized hydraulic parameters. Table 58 reports the scores attributed to each considered sensor (third phase) from the sum of which the final score (last column) is obtained (multiplying each score for the weight of the relative sensor specified in). The sensors that showed enhancements of the performance (positive score) with respect to the base simulation (VGM or HVGM) are underlined in pink in order to help the modeller in the visualization of the results. As could be observed, just a group of sensors could be optimized significantly using inverse analysis (scores >>0):MPS6 MPC1 3,05m, MPS6 MPC1 4,6m, GS3 MPB2 2,2m, GS3 MPC1 4,5m and GS3 MPC3 6,4m; while sensors PZ1 10m, T8 TC1 4,7m and GS3 MPC1 2,4m could not be optimized in any indirect simulation. PZ2 17m, MPS6 MPC3 6,2m, MPS6 SPC1 7,0m and GS3 SPC2 7,1m show just a limited optimization of the performance with respect to the simulation that uses the initial set of parameters and often this enhancement is not visible in a graphical way (the curve simulated dataset using the optimized parameters in time coincides with the curve simulated dataset using the set of initial parameters in time).

SIMULATION IDENTIF. CODE	1ST CALIBRATION PHASE	2ND CALIBRATION PHASE	3RD CALIBRATION PHASE
1_2_2	10,10	4,75	147
1_2_6	9,73	4,25	162
1_2_12	9,73	4,67	215
1_4_6	9,85	4,00	171
1_4_7	9,14	4,50	94
1_11_3	9,14	4,25	46
1_11_4	9,21	5,25	78
2_1_3	9,23	5,00	165
2_3_3	9,23	4,00	167
sim.2	9,73	5,00	152
MAX SCORE	14	8	
THRESHOLD SCORE	7,5	4	

Table 57: The table reports the scores of the 1st, 2nd and 3rd phase of the performance evaluation for each considered indirect simulation. The last row reports the threshold value set for each phase. In red the indirect simulations that do not overcome the validation phase.

	PZ1 10m	MPS6 MPB2 2,7m	PZ2 17m	T8 TB1 4,7m	MPS6 MPC1 3,05m	MPS6 MPC1 4,6m	T8 TC1 4,7m	MPS6 MPC3 6,2m	MPS6 SPC1 7,0m	T8 TC2 8,0m	GS3 MPB2 2,2m	GS3 MPC1 2,4m	GS3 MPC1 4,5m	GS3 MPC3 6,4m	GS3 SPC2 7,1m	FINAL SCORE
1_2_2	-4,3	5,8	0,0	-1,1	45,2	125,7	-4,2	7,5	3,6	-11,1	21,6	-152,7	27,2	68,7	5,5	147
1_2_6	-4,9	6,6	0,0	-1,0	45,8	127,4	-4,5	4,4	3,0	-7,4	19,8	-141,2	27,9	75,1	4,9	162
1_2_12	**-4,3**	**3,7**	**0,0**	**-1,0**	**46,4**	**128,8**	**-4,8**	**6,8**	**3,4**	**-10,6**	**6,0**	**-60,0**	**25,4**	**83,4**	**4,5**	215
1_4_6	-5,4	7,7	0,0	-0,9	46,5	129,4	-5,0	2,1	2,0	-3,4	17,5	-134,7	28,2	78,2	4,5	171
1_4_7	-3,4	-1,7	0,0	0,1	27,3	82,1	-0,1	-10,4	-3,0	4,0	-11,2	-97,8	27,7	71,8	5,9	94
1_11_3	-3,8	-202,3	0,0	1,2	45,2	125,5	-4,1	6,2	3,2	-7,7	162,3	-153,4	27,1	66,3	5,8	46
1_11_4	1,3	-1,4	0,0	0,1	-6,4	-28,7	-1,0	1,1	-0,5	0,7	176,7	-120,2	27,0	60,0	7,4	78
2_1_3	-3,1	4,2	0,0	-0,7	34,4	97,3	-0,9	7,2	3,2	-7,2	14,6	-71,9	21,7	72,2	2,6	165
2_3_3	**-3,2**	**4,3**	**0,0**	**-0,7**	**35,1**	**99,1**	**-1,0**	**7,3**	**3,2**	**-7,3**	**15,1**	**-75,0**	**22,3**	**73,7**	**2,8**	167
sim.2	-4,3	5,7	0,1	-1,1	44,9	124,9	-4,0	7,8	3,6	-10,6	21,7	-148,2	27,5	70,8	5,3	152

Table 58: The table reports in detail the scores attributed to each sensor in the third phase of evaluation of the indirect simulations performance (validation process). The scores of the sensors that result enhanced with respect to the base simulation are underlined in pink (values >0). In red the final scores of the indirect simulations whose performance is not sufficient.

Simulations 1_2_12 and 2_3_3 have shown excellent performances among the 10 considered ones (see Table 58) and are presented graphically in Figure 38 and Figure 39 by way of example. 1_2_12 enhances the performance of the HVGM while 2_3_3 of the VGM.

In Figure 38 and Figure 39, the simulated behaviour of each sensor (red line) is presented together with the behaviour of the base simulation (blue line) and of the observed dataset (green line) in the same observation point. As seen in the calibration phase, the enhancement of the behaviour of the considered sensors using the optimized sets of parameters is observable but not of great magnitude for the majority of the sensors. The same information was observed numerically in Table 58. The initial set of hydraulic parameters ensures right from the start a good initial performance of the model because proposed after an in-depth and large campaign of laboratory tests to investigate the retention properties of the soils.

The optimization process by means of inverse analysis is able to enhance the model performance guiding the modeller in the discovery of a better set/s of parameters but enhancements result localized to certain layers (sensors) or certain simulation periods.

We can conclude that globally the performance of the considered sets of optimized parameters in the validation phase is acceptable (excluding simulations 1_4_7, 1_11_3 and 1_11_4) because the simulated datasets follow the peaks of the observed datasets without delays (good simulation of the flooding events as desired). As seen in the calibration phase, it is not possible to obtain a perfect overlap of the simulated-observed datasets in the initial point due to the fact that the model uses as initial conditions the pwp distribution of the final time lapse of the 2018 period which shows some differences with respect to the observed pwp values. Hydrus does not allow to set in a discrete number of points values of water content or pressure head and perform an interpolation in the remaining nodes of the mesh to generate the initial conditions. The influence of the initial conditions during this one-year validation period seems of considerable relevance, since the floods events of great magnitude are registered only in the last half of the year and important floods are able to ''erase'' partially initial conditions influence. It is possible to state that in some cases, the influence of the initial conditions seems greater than the ''influence'' of the set of optimized parameters.

Table 59 reports the final sets of hydraulic parameters that have proven to optimize the performance of the model in the whole simulation period (from 2017 to 2019). In red the optimized hydraulic parameters that are different from the initial guess. As it is possible to observe, the parameters of the embankment layer

are the only ones to be optimized among the remaining indirect simulations, at the end of the validation phase. The α parameter (1/m) of the embankment layer seems moving toward higher values (lower air entry value) with respect to the initial guess (0,850 1/m). From this, we can understand that the behaviour of the embankment layer seems more influenced by the sandy fraction of the soil and less by the clayey fraction. The optimized saturated permeability Ks seems moving within an interval close to the initial parameter ($9*10^{-5}$ m/min), the residual water content θr (0,044) shows a change of an order of magnitude with respect to the initial guess (0,04) but as seen in the sensitivity analysis (chapter §5_1) it is the less sensible compared to the other hydraulic parameters. On the contrary, the hydraulic parameters Ks and α of the embankment layer are among the most sensible ones as showed in the sensitivity analysis of the hydraulic parameters performed in chapter §5_1. In Figure 37 the optimized parameters (blue dots) are presented on a normal distribution together with the initial guess of the parameter (purple dot) and the laboratory parameters (red dots).

All the remaining 7 indirect simulations could be considered possible final sets of optimized parameters and the scores proposed in the third phase to evaluate globally the performance of the simulation (last column in Table 58) and the scores attributed to each sensor (other columns in Table 58), could guide the modeller in the choice of the set of hydraulic parameters to use. Some indirect simulations behave almost identically (as 1_2_2 and 1_2_6) other enhance or worsen the performance of the considered sensors in way slightly different (as simulation 2 and 1_2_12). Based on the purposes/goals that the model wants to achieve, this choice could be made responsibly by the modeller in the light of the information obtained from the investigation performed on each indirect simulation. In the present case study, indirect simulations are excluded in the third phase based only on the final score because all the simulations optimize the parameters of the embankment layer whose hydraulic behaviour is the focus of the whole calibration and validation procedure. However if the purpose of the modeller is different (optimization of the performance of a certain layer, of a certain observation point or of a specific simulation period) the criterion of exclusion could be based not only on the final score (f.e. the choice falls on the set of parameters that enhance the layer of our interest even if it doesn't have the best score).

	embankment									fluvial								
	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw
1_2_2_HVGM	0,004	0,397	1,704	1,304	9,00E-05	0,398	0,397	3,408	9,00E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_2_6_HVGM	0,004	0,397	1,640	1,304	7,49E-05	0,398	0,397	3,280	7,49E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_2_12_HVGM	0,044	0,397	1,830	1,304	7,89E-05	0,398	0,397	3,660	7,89E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
1_4_6_HVGM	0,004	0,397	1,609	1,304	6,26E-05	0,398	0,397	3,218	6,26E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
2_1_3_VGM	0,004	0,397	1,239	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
2_3_3_VGM	0,004	0,397	1,223	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
simulaz 2	0,004	0,397	1,67	1,304	9,00E-05					0,0005	0,319	0,251	1,284	2,76E-05				
initial parameters	0,004	0,397	0,850	1,304	9,00E-05	0,398	0,397	1,700	9,00E-05	0,0005	0,319	0,251	1,284	2,76E-05	0,319	0,319	0,502	2,76E-05
	foundation									aquifer								
	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw	θ r	θ s	α	n	Ks	θ m	θsw	αw	Ksw
1_2_2_HVGM	0,001	0,424	0,12	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_2_6_HVGM	0,001	0,424	0,12	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_2_12_HVGM	0,001	0,424	0,12	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
1_4_6_HVGM	0,001	0,424	0,12	1,142	1,98E-07	0,424	0,424	0,236	1,98E-07	0,010	0,43	0,200	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05
2_1_3_VGM	0,001	0,424	0,12	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
2_3_3_VGM	0,001	0,424	0,12	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
simulaz 2	0,001	0,424	0,12	1,142	1,98E-07					0,010	0,43	0,200	1,200	9,00E-05				
initial parameters	0,001	0,424	0,12	1,142	1,98E-07	0,424	0,424	0,236	1,980E-07	0,010	0,43	0,2	1,200	9,00E-05	0,43	0,387	0,400	9,00E-05

Table 59: The table reports the sets of optimized parameters of the 7 indirect simulations that have overcome the validation phase. In red the hydraulic parameters different from the initial guess. The parameters α and αw are in 1/m, Ks and Ksw in m/min, θs, θr, θm, θsw, have no physical dimensions.

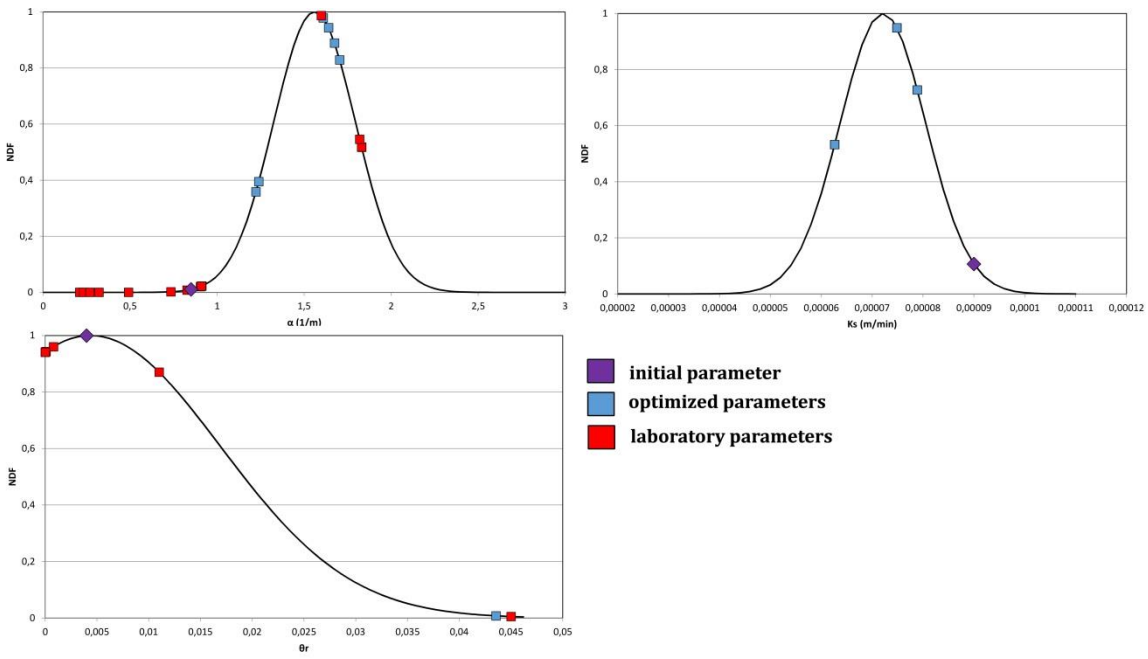


Figure 37: Normal distribution functions of the optimized hydraulic parameters (blue dots), of the laboratory parameters (red dots) and of the initial parameter (purple dots).

8_1 PRELIMINARY CONCLUSIONS AND FURTHER STUDIES

The last phase of the optimization of the model parameters is the model validation. The sets of optimized parameters that overcome the three phases of the calibration protocol in the previous chapters have been tested on a new monitoring period (24th June 2018-30th June 2019) using a new set of observed dataset. The 2019 period is characterized by 4 months of low river water level and 7 months of high water level which allows testing the models in both the hydraulic conditions. No changes to boundary conditions and geometry have been introduced with respect to the previous model of the 2018 period. As initial conditions, the pwp distribution of the last instant of the 2018 simulation periods has been imported.

The same procedure in three phases used for the calibration phase is replicated in the validation phase in order to test the predictive ability of the models. In total, the performances of 10 indirect simulations have been investigated. The indirect simulations show a good performance in the flood peak periods and a poor performance in the off-peak periods; so globally the performance of the model in the 2019 year is acceptable being halfway between the two trends. It has to be remembered that for the purposes of our modelling the model performance during the flood peak periods result of higher relevance with respect to the off-flood periods. The third phase of the performance evaluation shows that the enhancements are localized in certain layers or limited to certain sensors and simulation periods, as observed also in the calibration phase. For this reason the choice of the best set/s of observation points has to be made based on the purposes of the modelling and not only on the score achieved in the evaluation process. More relevance could be given to particular sensor (observation point), to certain hydraulic parameters for their higher sensitivity or the relevance of the layer in which they are installed or to certain simulation periods more interesting for the final goal of the modelling. In the presented case study, all the sets of metrics that overcome the three phases of the validation process (7/10) optimize the parameters of the embankment layer and the simulation of its hydraulic and retention behaviour is the focus of the entire modelling so the maximum score has been used as unique parameter of choice of the final sets of parameters. It must be

emphasize that the model output seems to depend greatly on the initial conditions that are imported from the last temporal instant of the previous model (2018 year) and a certain gap between observed and simulated pwp distribution is observable at time zero. This influence on the trend of the observation points seems to not decrease in time due to the fact that the initial part of the 2019 simulation period does not present any flooding events able to erase or partly erase the influence of the chosen initial conditions. Hydrus2D allows to set initial conditions in a simplified manner, while importing the observed values of pwp in the measurement points and interpolating in the remaining nodes could give better results. Another relevant observation is that the 7 successful sets of optimized parameters do not give great enhancements of the model performance with respect to the base set of hydraulic parameters (average values of the laboratory parameters). The same was observed in the calibration phase on the 2018 simulation period. This is due to the fact that the initial set of hydraulic parameters assures a sufficient/good performance of the models and further enhancements are difficult to obtain expecially in such a complex and multi-layered model.

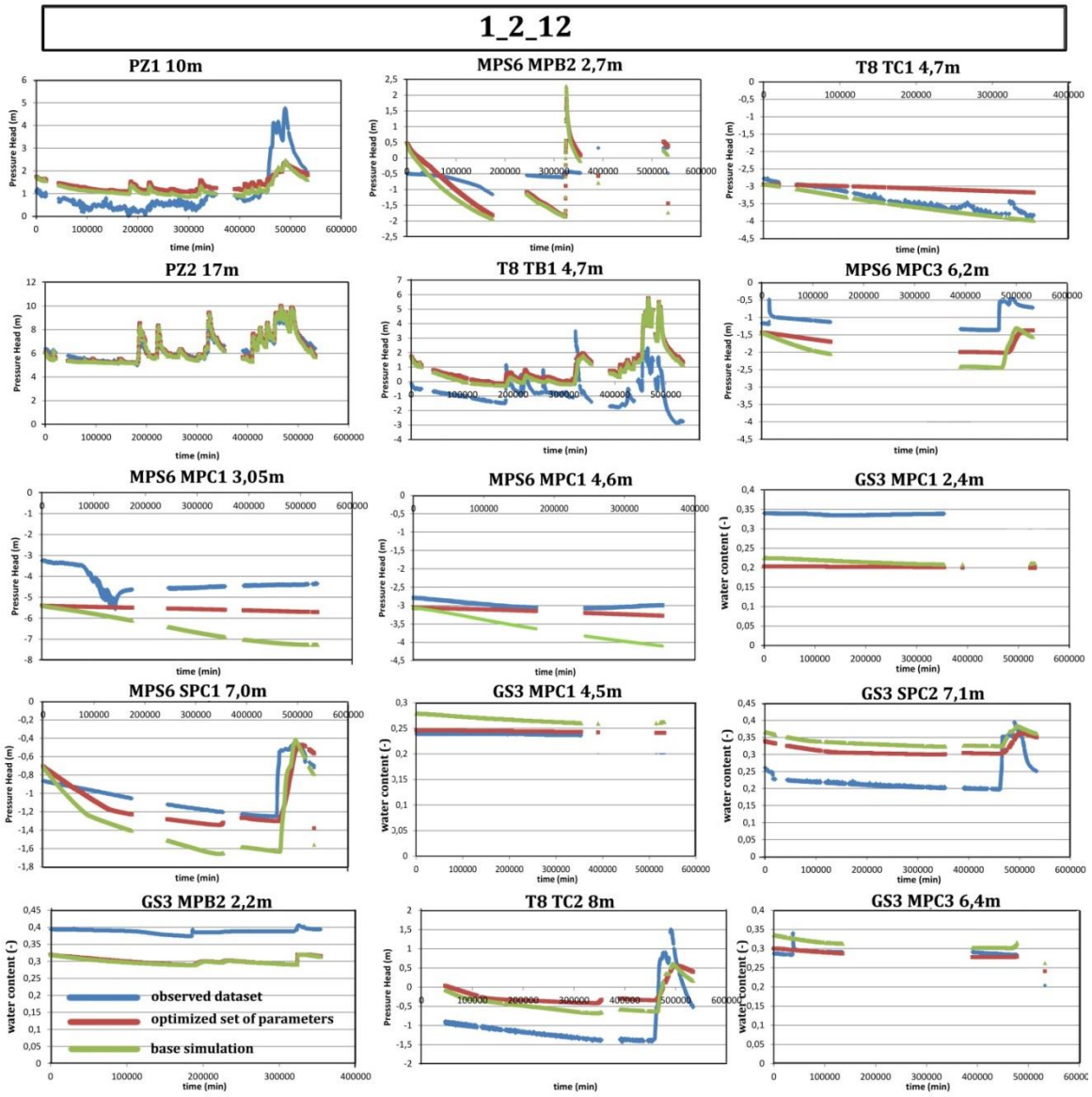


Figure 38: The graphs present the observed dataset (blue line), the base simulation (that uses the set of initial parameters) (green line) and the optimized dataset (parameters of inverse analysis 1_2_12) (red line) vs time, in order to allow a graphical comparison.

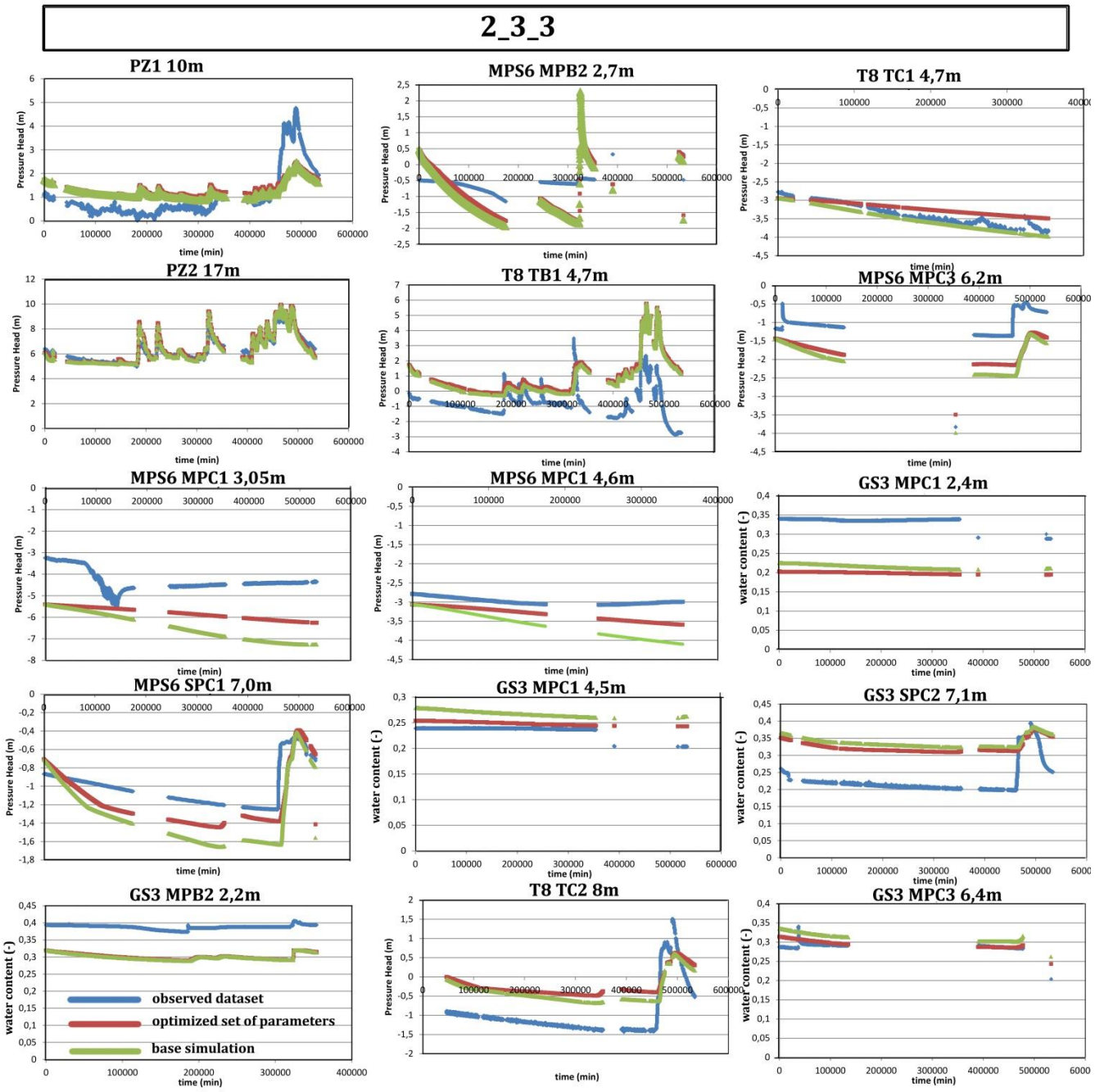


Figure 39: The graphs present the observed dataset (blue line), the base simulation (that uses the set of initial parameters) (green line) and the optimized dataset (parameters of inverse analysis 2_3_3) (red line) vs time, in order to allow a graphical comparison.

9 THE SCALAR FIELD OF THE FACTOR OF SAFETY (SFFS)

The limit equilibrium methods are considered the most effective and reliable to perform slope-stability analyses. The main limits are related to the geometry and the position of the potential failure surface which is sometimes predetermined or approximated. The 2D limit equilibrium methods are based on the discretization of the potential failure surface in vertical slices considered as sliding blocks. The main limitation is that the geometry of the instable mass is assumed and the failure occurs simultaneously in every point of the sliding surface. The main goal is to find a global indicator of the slope stability called FS, Factor of Safety, defined as the ratio of the soil shear strength to the soil shear stress needed to ensure equilibrium along the sliding surface. A number of different variations of the limit equilibrium method are proposed depending on how the inter-slice forces are considered, the equilibrium equations considered and the shape of the failure surfaces (e.g. Janbu, 1954; Bishop, 1955); Morgenstern and Price (1967); Spencer (1967). All these methods do not allow to have insight into the location of the initial failure surface and predict the local change of the potential instable mass due to external stimulus (river water height or rainfall infiltration). To track how the potential failure surface initiates and evolves, the global Factor of Safety is not helpful.
To overcome these limitations a scalar quantity that could be computed in each point of the hillslope could be used (Iverson and Reid, 1992; Lu et al, 2012): the Scalar Field of the Local Factor of Safety (SFLFS). To define this indicator, the current state of stress in each point τ is divided for the shear strength τ* obtained shifting the Mohr circle representing the actual stress state leftward until it touches the M-C envelope criterion (see `Figure 40`). In fact the changes of the state of stress due to pore water pressure variation in an unsaturated embankment cause a similar leftward shift of the Mohr circle (Lu et al, 2012). This is due to the fact that infiltration and seepage phenomena cause a reduction in the size of the Mohr circle (given by the difference between the major principal effective stress σ'_1 and the minor principal effective stress σ'_3) of less extent than the variation of the suction stress which controls the position of centre of the circle.
As already mentioned, the Local Factor of Safety is defined as follows [equation 146]:

$$\mathrm{LFS} = \frac{\tau}{\tau*} \qquad \text{[equation 146]}$$

Where τ is the current Coulomb stress and τ* the potential Coulomb stress. LFS is a scalar quantity that changes in time and space. LFS is equal to 1 if the stress state is equal to the limit state and the potential local failure could occur.
The LFS includes both the extensional and compressive stress failure regimes. ''Slope Cube'' is an add-on module of Hydrus2D and it has been chosen over other commercial programs because it interfaces easily with the simulations performed, with no compatibility issues between the Hydrus output and the computation of the LFS using the imported pwp distributions.

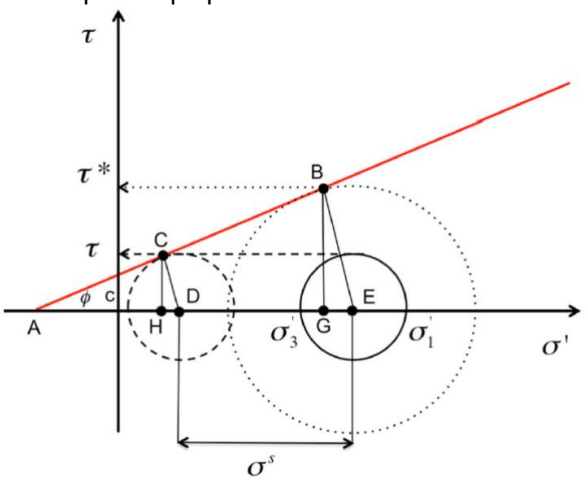


Figure 40: Conceptual representation of LSF (from Lu et al, 2019)

The Slope Cube Module couples variably saturated flow solving the Richards equation and the total stress field obtained solving the elasticity problem (Lu et al, 2012). The module is able to compute effective stress, deformations and the Local Factor of Safety in each point of the whole domain. The effective stress model adopted in Slope Cube module was proposed by Lu and Likos (2004 and 2006). It has the same form of Terzaghi's effective stress for saturated soils [equation 147]:

$$б' = б - б^s \quad \text{[equation 147]}$$

Where б' is the effective stress, б is the total stress and $б^s$ is the suction stress.
Suction stress is defined trough the van Genuchten's retention model as follows [equation 148]:

$$б^S = -\frac{(u_a - u_w)}{(1+[\alpha(u_a - u_w)]^n)^{(n-1)/n}} \quad \text{[equation 148]}$$

where $u_a - u_w$ is the matric suction, and α and n the VGM parameters. To compute the strain field, the Hooke's law is used, no plastic deformations development is considered during the slope failure. The mechanical model has implemented a variable elastic modulus E depending on the volumetric water content. The Young moduli E(θ) is computed as follow (Lu and Kaya, 2013) [equation 149]:

$$E(\theta) = E_d + (E_w - E_d)(\frac{\theta - \theta_d}{\theta_w - \theta_d})^m \quad \text{[equation 149]}$$

where E_d is the Young moduli at the residual water content, E_w at saturation, m is an empirical parameter, θ the water content, θ_w is the saturated water content and θ_d the residual water content.

Figure 40 from Lu et al (2012) shows τ and τ* on the Mohr-Coulomb space, where $б'_1$ and $б'_3$ are the major and minor principal effective stress respectively. Using the triangle properties, LFS could be rewritten in terms of the mechanical parameters (c', θ') and of the total stresses $б_1$, $б_3$ and suction stress $б^s$ as follow [equation 150]:

$$LFS = F(x,y,z,t) = \frac{cos\phi'}{6_1 - 6_3}[2c' + (6_1 + 6_3 - 26^2)\, tan\phi'] \quad \text{[equation 150]}$$

LFS could vary in the positive and negative range depending on the values assumed by $б_1$, $б_3$ and $б^s$. Values of LSF below 0 suggest possible instable masses and creation of possible sliding surfaces following their evolution over time. The field of LFS could be used as a quick and initial check of the most relevant time steps of the simulation in order to understand when it is necessary to perform more sophisticated stability analyses.

It is relevant at this stage to investigate the consequences of using the optimized parameters datasets instead of the initial parameters dataset on the pore pressure distribution during the 2018 simulation period. In Figure 41, the distribution of the Local Factor of Safety is proposed in three temporal instants 40600 min (10/12/2017, flood C in Figure 19), 175000 min (13/03/2018, flood G in Figure 19) and 210000 min (06/04/2018, flood H in Figure 19) of the simulation period, corresponding to three very significant flooding events. The chosen optimized dataset is 1_2_12 (HVGM) whose parameters are reported in Table 60. In Table 61 the mechanical parameters used to compute the field of the LFS are reported for each layer of the riverbank section under investigation. The young moduli Ex (horizontal direction) and Ez (vertical direction) are computed from the in situ cone penetration tests (CPTU) and laboratory tests (oedometer tests) when available.
The isolines represent the boundaries of regions which have the same range of LFS. LFS is considered ranging from 0 (blu) to 1 (red) in Figure 41. LFS equal to 1 means that there is no reserve of resistance (τ=τ*) while LFS ≤0 indicates a soil mass prone to collapse.

In Figure 41 the cross sections obtained using the initial parameters dataset (1_2_12) and the base dataset are presented. The differences between the sections using the base dataset and the optimized dataset are observable but not of great magnitude as the differences between the optimized hydraulic parameters and the set of initial parameters and these differences mainly concern the embankment layer. Four vertical sections (n°1, 2, 3, 4) passing through the considered riverbank are represented in the last image in Figure 41: two in the berm, one in the inner slope and one in the crest.
In the graphs in Figure 42, the variation (%) of the Local Factor of Safety in the three different bank sections is presented in three different time instants (40600, 175000, 210000 min). The variation of the LFS is used to quantify the differences between the simulation using the base set and the optimized set of parameters. The percentage variation is computed as follows [equation 151]:

$$\text{LFS variation } (\%) = \frac{(LFS_{optimized} - LFS\ initial)}{LFS\ initial} * 100 \qquad \text{[equation 151]}$$

where ''LFS optimized'' is obtained using the set of optimized parameters, ''LFS initial'' using the set of initial parameters. As it is possible to observe between the soil surface and 10 m of depth the variation (%) ranges between 3-5%, between 10 m - 20 m the variation reaches up to 22% then its value drops again. The major differences are observed for the temporal instant 175 000min (13/3/18) during flood G (Figure 19) the more relevant of the 2018 year for river water level reached and persistence in time and for the vertical sections 3 and 4 located in the berm.
The main purpose of the present thesis lies outside the topic of embankment and slopes stability analysis, but it concerns a previous level of definition of the pore-pressure distribution to be imported for stability analysis. Figure 41 and Figure 42 want to emphasize the variation in the values of the Local Factor of Safety as consequences of the pore-pressure distribution variations using the initial parameters dataset and the optimized parameters dataset as obtained with the present calibration/validation procedure.

HYDRAULIC PARAMETERS- INDIRECT SIMULATION 1_2_12									
	θ_r (-)	θ_s (-)	α (1/m)	n (-)	K_s (m/min)	θ_m (-)	θ_{sw} (-)	α_w (1/m)	K_{sw} (m/min)
embankment	0,04354	0,3970	1,830	1,304	7,886E-05	0,3981	0,397	3,660	7,886E-05

Table 60: Hydraulic parameters of the indirect simulation 1_2_12.

PARAMETERS – SLOPE CUBE SIMULATIONS						
LAYERS	**Gs (-)**	**c' (kPa)**	**ϕ' (°)**	**Ex (kPa)**	**Ez (kPa)**	**v (-)**
FLUVIAL	2,523	0	33,4	6300	6300	0,33
FOUNDATION	2,530	0	24,9	9800	9800	0,33
AGRARIAN SOIL	2,646	0	30,6	7700	7700	0,33
AQUIFER	2,500	0	31,5	14200	14200	0,33
SUBSOIL	2,500	0	22,5	12300	12300	0,33
EMBANKMENT	2,646	0	30,6	6600	6600	0,33

Table 61: Mechanical and stiffness parameters of the layers under investigation.

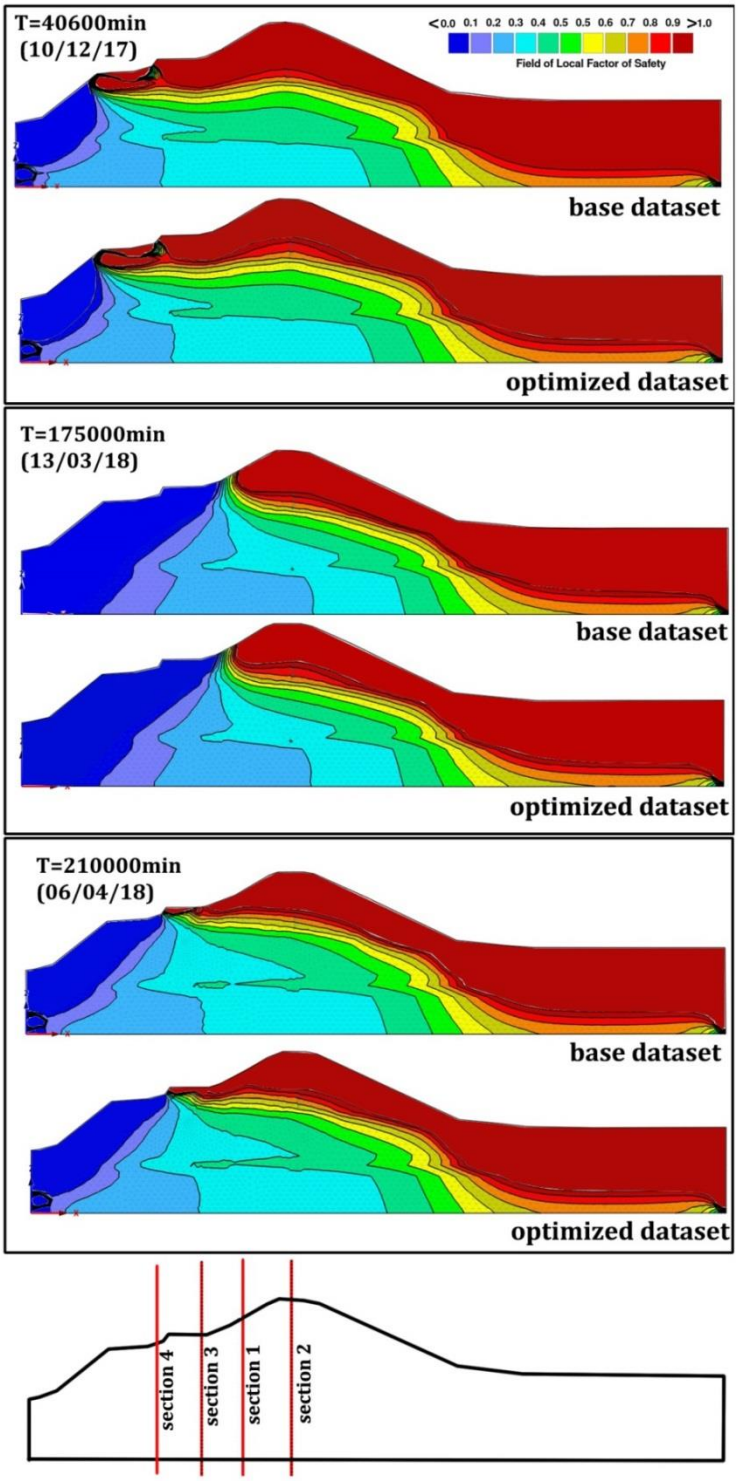


Figure 41: Representations of the Field of the Local Factor of Safety (FLFS) for the investigated riverbank section in three temporal instants (40600 min, 175000 min and 210000 min) using the pwp distribution modelled with the set of base parameters and the set of optimized parameters.

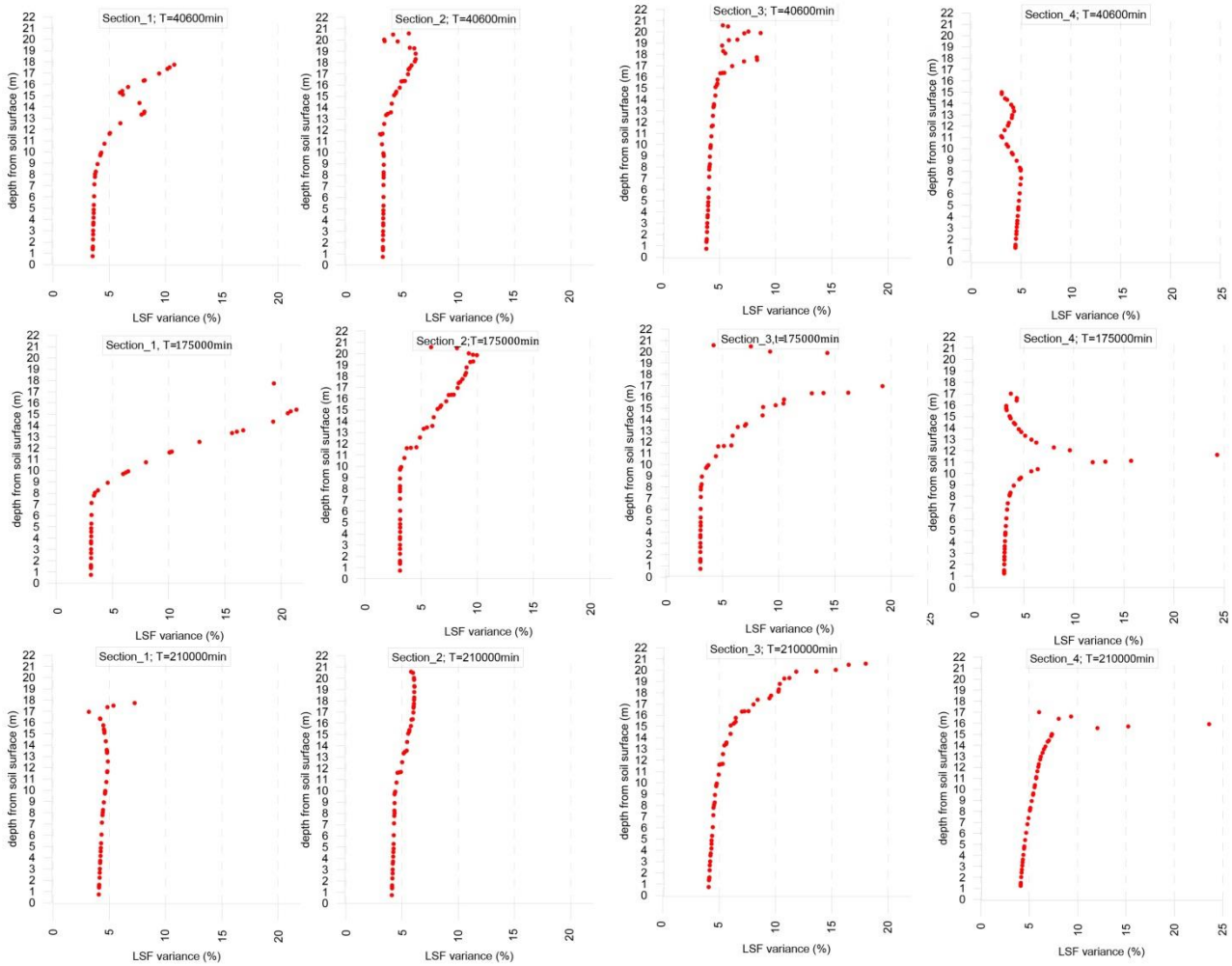


Figure 42: Variance in percentage between the Local Factor of Safety computed using a pwp distribution simulated with the set of base parameters and the one simulated with the optimized set of parameters (1_2_12) in three relevant temporal instants (40600 min, 175000 min, 210000 min).

10 CONCLUSIONS PART I

The present work guides the modeller step by step in the quantification of the performance of a model output. Most of the times, the results of a calibration/validation procedure are evaluated based on subjective decisions/personal evaluations of the modeller but this method lacks in transparency and need a great experience in modelling. The focus of the present work is on the simulation of the hydraulic and retention behaviour of a riverbank section subjected to transient seepage conditions. The chosen case study is a riverbank section along Secchia River, close to Cavezzo (MO) and the simulation period is the 2018 year (12th November 2017 to 23rd June 2018). The proposed procedure of model performance evaluation could be applied to:

- direct simulations that model the soil-water dynamics within an embankment section using different hydraulic models in order to individuate the one/s that better represent the observed behaviour.

- direct simulations in the calibration phase, when a trial and error procedure is applied in order to individuate the combination of parameters which enhance the model performance.

- indirect simulations (inverse analysis) carried out in order to optimize the model performance (reduction of the deviation between observed-simulated datasets) individuating the set of hydraulic parameters which better represent the phenomenon observations.

To perform inverse analysis, a dataset of observation points is needed (data from in situ acquisitions or laboratory tests). For the presented case study, the observed dataset consists of in situ acquisition of water potential and water content by sensors installed in the investigated bank section (see subchapter §3_1). The great strength of the inverse analysis programme performed stays in the large availability of in situ data in different positions of the bank section and at different depths from the soil surface, even at great depth (see subchapter §3_1).

Four different hydraulic models have been used to model the behaviour of the riverbank section: the van Genuchten Model (VGM) (6 hydraulic parameters to describe the main drying curve), the van Genuchten Model considering hysteresis (HVGM) (10 hydraulic parameters, 6 to describe the main drying curve and 4 to describe the main wetting curve), the VGM and HVGM considering an air-entry value of -2cm (VGM-2cm and HVGM-2cm, respectively) in order to force the slope of the retention curve to be zero close to saturation and obtaining, as a result, that the hydraulic conductivity function is less non-linear and more accurate close to saturation (for more information see chapters §2_3, §2_4 and §2_5).

Metrics and indices have been tested prior to their application to the simulation output using synthetic datasets of type water pressure head, water content and both in order to guarantee that the trend of the metrics/indices is always consistent with the expected logical directions without ambiguous performances (see chapter §4). The accuracy metrics MAE, MSE, RMSE have shown for all the considered cases of observed datasets logical trends, R^2 has shown in some cases illogical fluctuation of its value due to the sensitivity to outliers and insensitivity to additive and proportional differences between observed-simulated datasets. Despite this, its use together with other metrics is always suggested because it is very well known and widely used in the field of the simulation performance evaluation. The efficiency metrics NSE, NSE_{rel} and NSE_j have shown in all cases consistent and logic behaviours despite their sensitivity to outliers while IA, IA_j and IA_{rel} are replaced by the modified form IA' which shows a good performance in all

the casuistries. COV, NCOV and KGE show a logical behaviour (the only exception is COV with a mixed dataset and KGE with a water content dataset). For more information on limits and potentialities of each metric/index used in the presented work see chapter §2_7_2. Due to the fact that each metric/index has its personal pros and cons, it is always suggested to evaluate the performance of the simulation using a set of different metrics (accuracy, efficiency, model selection criteria) in order to have a more complete overview on the goodness of the simulation output, less affected by the limits of the single metrics. Despite this, the use of a large number of metrics/indices is not suggested because the modeller could be lost in the high number of values to process resulting unable to draw any conclusions. A balanced number of accuracy metrics, efficiency metrics and model selection criteria on which the modeller has confidence is the key for an optimal model evaluation.

Chapter 5 proposes an in-depth sensitivity analysis of the hydraulic parameters of the main wetting curve and of the main drying curve in order to individuate the parameters whose changes influence the most the model output. To do so, a sensitivity coefficient is used to compare objectively the output of the simulation using the base parameter and the output using the varied value (variation of +/-1% with respect to the initial value). One single parameter at at time is changed (OFAT method). It has been observed that n, θs of the embankment layer show the highest sensitivity in correspondence of two of the main peaks of the 2018 simulation while in the remaining period the sensitivity is low or close to zero. For what concerns the fluvial layer, n, θs, α show an high sensitivity which is zero in the rest of the simulation period. As expected the saturated permeability of the aquifer layer is the only sensitive hydraulic parameter. In the foundation layer, a significative sensitivity is registered for all the hydraulic parameters during the whole investigated period, expecially the n parameter. The parameters of the wetting curve have a lower sensitivity compared to the parameters of the drying curve and peaks are observed for the fluvial and embankments layers in correspondence of the two main flood peaks, while for the foundation layer the sensitivity is more homogeneous over the simulation period.

Prior to the performing of the calibration procedure, the modeller needs to select the hydraulic model that best represents the simulation phenomena in order to reduce the number of indirect simulation to be carried out, focusing the attention on a single model. It is important to understand if increasing the complexity of a model passing from a lower number to a greater number of parameters is meaningful and desirable based on the purposes of the simulation that the modeller is performing. Moreover it is always important to understand which is the starting point of the calibration process, in other world, the goodness of the performance of the base simulation in order to understand the entity of the improvement to aim for in the calibration phase.

In the presented case study, however, all the 4 hydraulic models have been optimized using inverse analysis because the main goal is to test the potentialities and limits of the Marquardt-Levenberg optimization algorithm implemented in the commercial code Hydrus2D by Pc-Progress. The performances of the four different hydraulic models applied to the simulation of the 2018 year have been investigated, initially using a graphical representation (comparison between the observed dataset and the simulated dataset) then using a set of chosen metrics/indices in order to give a quantification of the models performance allowing an easy comparison between them. For the presented case study, VGM and HVGM show the best performance with values of the metrics close to the best fit, HVGM is able to simulate better the major flood events while in the remaining periods the differences between the two models are low or close to zero (see chapter §6_1). The HVGM is able to reproduce with more precision the retention behaviour in the fluvial layer (berm, see Figure 9) because the hydraulic stimuli affect earlier this area and

here they are more persistent in time, showing the importance of considering the soil hysteretic behaviour. VGM-2cm shows a performance slightly poorer compared to the VGM and HVGM ones. HVGM-2cm shows the poorest performance when compared to the observed dataset. Due to the fact that the focus of the calibration programme is the enhancement in the simulation of the pwp distribution in the major flood events of the 2018 year, it is always suggested to match the use of the set of metrics on the whole simulation period to the use of the same metrics/indices on the major flood events and following drawdowns separately. This allows to understand where the main limits in the model simulation stay in order to focus the calibration programme on their resolution and in order to limit the number of inverse analysis to be performed. Moreover, it is important to remember that the flood events with medium/high river water level but long persistence in time are the most relevant for fine-grained riverbanks (as the one considered in the case study) because the saturated profile has time to propagate deeper in the bank body. For this reason a particular attention must be paid to these simulation periods. Differently, if the riverbank section under investigation is in coarser material, the modeller must focus the attention on the flood events with highest river water level regardless of their persistence in time. For the presented case study, VGM and HVGM show the best performance in the flood peak periods while a poorer (but good) performance during the off-peak periods. The performance of the four hydraulic models has been investigated using three different typologies of observation dataset (water pressure head, water content, pressure head+water content). For all the considered hydraulic models the best performance has been shown using a pressure head dataset, which is extremely positive if we consider that reliable pwp distributions in the most interesting time instants of the simulation are the final goal to achieve. The poorer performance has been shown by the water content dataset while the mixed dataset (water content+pressure head) has a performance in between.

A large campaign of inverse analysis has been performed: 1028 indirect simulations optimizing different combination of hydraulic parameters (from 1 to 15 parameters) belonging to one layer up to 3 layers together, using different measurements type (pressure head, water content or both) and different weighting distributions (triangular moving average technique and all observed data with the same weight equal to 1). Despite a sensitivity analysis has been performed on the parameters of all the considered hydraulic models (see chapter §6_1), the indirect simulations try to optimize all of them without exclusion for low or very low sensitivities. This choice was dictated by the necessity to conduct a very ‘‘general’’ inverse analysis programme able to deal with the largest variety of problems encountered in the calibration of a complex model, in order to be applicable to other similar case studies.

The 25% of the performed indirect simulations reaches convergence and, with high probability, the high percentage of unconverged simulations is due to the complexity of the problem (multi-layered domain, medium-long simulation period, transient seepage conditions, important flooding events). The VGM shows the highest percentage of successful simulations (46%) with respect to the initial number, HVGM the 41% , VGM-2cm the 14% and HVGM-2cm the 1%. The HVGM-2cm shows a percentage of successful simulations close to zero due to ill-posedness issues of the indirect problem. It is impossible to know a-priori if an indirect problem is well or ill-posed and there is not an univocal way to solve the problem: the optimization of a small number of hydraulic parameters and the use of a large and reliable observation dataset are valuable suggestions to overcome the problem. Both the strategies have been tried but with no success. The low number of successful simulations for the HVGM-2cm makes impossible the application of the calibration protocol and for this reason the simulations are excluded from the following phases. The embankment and foundation layers (VGM, HVGM) have the highest percentage of successful simulation

(>50%) with respect to the performed initial number, while the fluvial layer has a lower percentage (<36%). All the considered hydraulic models show the highest percentage of successful simulations when a pressure head observed dataset is used, the lowest percentage for a water content dataset and for the mixed one (PH+wc) a percentage in between. The Triangular Moving Average Technique (TMAT) shows a lower percentage of successful simulations with respect to the weighting distribution equal to unity for all the observation points. This is due to the fact that the TMAT gives more weight to the flood events (with respect to the off-flood periods) and the optimization of the set of hydraulic parameters results more complex when the focus is on the reduction of the deviation between observed-simulated datasets in the flood peaks. Moreover the quality of the match observed-simulated in the flood periods is high and a further enhancement does not come easily.

A procedure of evaluation of the indirect simulations performance in three phases is adopted in the calibration process in order to investigate and compare the simulations in a quantitative and objective way.

A pre-calibration phase is used when a large number of indirect simulations is available whose performances must be evaluated quantitatively. A set of metrics/indices is used for this purpose (accuracy, efficiency and model selection criteria) and for each of them a threshold value is imposed. The metrics are applied to the whole simulation period in order to have an overall idea of the simulations behaviour without going into details (the proper analysis is left to the next calibration phases). No scoring system is set in the pre-calibration phase but only the overcoming (or not) of the threshold values imposed for each metric/index is observed. This pre-phase allows to decrease the number of inverse analysis excluding the ones that behave poorly and focus the attention on a restricted number. For more information see chapter §7_2.

The first phase of the calibration is based on a scoring system: points are given according to the ability of each metric to predict the performance of a simulation and according to the distance from the best fit. The focus of the analysis is the deviation between observed-simulated datasets. The maximum score which can be attributed to each simulation (15) is subdivided in 7,13 points for the performance of each simulation in the peak periods, 2,38 points for the off-peak periods and the remaining points for the performance on the whole simulation period. The points in the peak and off-peak periods are subdivided according to the river water height reached and the persistence in time. The threshold score 7,5 over 15 sets the overcoming of the first phase and the transition to the second phase of the calibration. The 80% of the initial number of indirect simulations has overcome the first calibration phase which is less ''selective'' with respect to the following calibration phases (second and third). The indirect simulations optimizing the embankment layer (for the VGM and HVGM) show the best performance (highest score) and the largest number of successful simulations that overcome the first calibration phase. This is due to the fact that a greater number of sensors has been installed in the embankment layer with respect to the other layers and the use of a larger observed dataset has beneficial influences on the results of the optimization. On the contrary the foundation layer (for the VGM and HVGM) shows the lowest percentage of successful simulations (<60%) due to a restricted number of available observation points in the layer. It has also been noted a low percentage of successful simulations for the optimization of groups of layers as embankment+foundation and embankment+foundation+fluvial, for the VGM and HVGM. This is due to the increasing difficulty in obtaining a good performance of the inverse analysis as the number of parameters and layers increase in the analysis. The VGM-2cm shows a general poor performance and a low percentage of successful simulations (that overcome the calibration phase) with respect to the VGM and HVGM, independently of

the weighting distribution adopted and which layer has been optimized. For more information see chapter §7_3.

The second phase of the calibration focuses on the results of the inverse analysis, looking closely to its last iteration and on the obtained optimized parameters in order to evaluate the goodness of them. Due to the fact that the purposes are completely different from the ones of the first calibration phase, the adopted set of indices is different. The 95% confidence limits, the mass balance error, the correlation coefficient of the parameter, the simulation tolerance, the number of optimized parameters, the number of iterations, the standard error coefficient, the difference between the sum of the square residual between the zero iteration (with the initial set of parameters) and the last iteration (with the optimized set of parameters) are considered. As for the first phase of the calibration, a scoring system is used based on the closeness to the best fit of each index. 8 is the max score that could be obtained by each simulation, 4 is the threshold value. Looking closely to the values of the adopted indices, it is possible to draw specific conclusions on the investigated indirect simulations. For example the mass error gives information on solution accuracy and uncertainties in the predictions, the number of iterations gives information on the closeness or not of the optimizes parameters to the initial set of parameters, the tolerance gives information on the precision with which the inverse problem has been solved, the correlation coefficient on the possibility to determine together couple of parameters suggesting when the determination of the parameters singularly has to be preferred. The 53% of the initial number of simulations (that passed the 1st calibration phase) has overcome also the second calibration phase which is more ''stringent'' and ''selective''. As observed in the first phase of the calibration, the PH dataset has the highest number of successful simulations (75% for the HVGM, 81% for the VGM) while the wc dataset a lower percentage (43% HVGM, 64% VGM). The weighting distribution TMAT confirms also in the second calibration phase its poor performance, allowing to conclude that its use in the inverse analysis does not guarantee a good performance of the indirect simulations. Even in the second calibration phase, the HVGM has the highest percentage of successful simulations with a score above the imposed threshold, while VGM-2cm has such a low percentage of successful simulations that it will be excluded from the following calibration phase. Increase in the magnitude of the residuals, problems of convergence and parameters correlation suggest the use of an inappropriate model as in the case of VGM-2cm. Any indirect simulation optimizing the hydraulic parameters of the wetting curve (HVGM) has overcome the second calibration phase but, as seen in chapter §5_1, the sensitivity of these parameters is extremely low with respect to the wetting curve parameters as a result of an hysteretic behaviour of the riverbank not very marked. It has been observed that a good score in the first phase (good match observed-simulated datasets) does not exclude problems of low accuracy and correlation of the parameters, this makes necessary and legitimate to test each indirect simulation in the second calibration phase. It has been observed that the majority of the simulations that overcome both the calibration phases, optimize a low number of parameters this is due to the fact that model uncertainties and problems of parameters identifiability reduce when a limited number of parameters is optimized. The use of a Response Surface Analysis has been proposed in order to investigate the behaviour of the objective function in specific parts of the parameters space when problems of parameters correlation, non-identifiability or the existence of multiple minima of the objective function are suspected. However, considering the computational effort and time consumption of this analysis, the modeller could refer to this type of analysis only in a limited number of cases when he needs to corroborate (or not) hypothesis based on the results of the calibration process (values of the indices). For more information see chapter §7_4.

The third phase of the calibration focuses on the performance of the single sensors on the whole simulation period and on the most relevant flood events of the 2018 year (two events in December 2017 and March 2018). To do so, the match between observed-simulated datasets is investigated in detail. In the first calibration phase the performance of the whole set of sensors has been investigated on the whole simulation period and on the flood peak and off-peak periods. Due to the fact that the number of indirect simulations is now reduced due to the previous calibration phases a more in-depth investigation of the behaviour of the single observation points could be carried out. A restricted number of metrics/indices (7) has been used (accuracy and efficiency metrics) in order to avoid to be lost in a large number of values to process for each sensor. The performance of each indirect simulation is compared to the one of the base simulation (that uses the set of initial parameters) that is considered as the reference performance. Scores are computed for each considered metric based on the enhancement with respect to the reference. Weights are given to the considered sensors based on their different position in the bank section and different typology of recorded information (water potential or water content). The same information could be obtained qualitatively in a graphical way plotting the simulated dataset of the optimized parameters, the base dataset and the observed dataset. It has been observed that any indirect simulation is able to enhance the performance of all the considered observation points. This results in the necessity to focus the attention of the modeller on the most relevant observation points/sensors for the purpose of the research. In our case, a reliable map of the pwp distribution in the most relevant time instants of the simulation is the final goal and it requires to concentrate the attention on the sensors monitoring pwp located in the unsaturated zone of the embankment layer. 9 indirect simulations have overcome the third calibration phase, all of them use a VGM and a weighting distribution equal to 1; 7/9 optimize the hydraulic parameters of the embankment layer and 2/9 the hydraulic parameters of the embankment+fluvial layers. It has been observed that the enhancement in the sensors performance are in all cases of small entity. This is due to the fact that all the performed indirect simulations start from a good set of hydraulic parameters obtained from a large campaign of laboratory tests (see chapter §3_2) that ensures a very good initial performance of the models. The achievement of an improvement in the model performance in these conditions is more difficult and requires a more meticulous work. Reliable initial parameters allow to start from a good performance of the model to save time in the calibration phase avoiding to perform a large number of inverse analysis. This is especially important when the model is complex (multi-layered section; multi-parametrical models) and parameters correlation, convergence issues and non-identifiability of certain parameters are relevant problems observed with greater incidence. The sets of optimized parameters could not be considered a statistical family because obtained under different conditions (different datasets, different weighting distributions, different combination of optimized parameters), moreover the initial number of performed indirect simulations is large but not enough to be considered statistically significant. A deterministic approach and not a probabilistic one is followed in the presented work that has the main purpose to individuate the set/s of hydraulic parameters that enhance the most the model performance among the discrete number of indirect simulations performed. Despite this, a graphical representation on a normal distribution of the optimized parameters allows to observe their dispersion with respect to the initial parameters and the laboratory parameters (hydraulic parameters obtained in the laboratory tests) and with respect to the type of observation dataset used (water pressure head or water content or both). It has been observed that in the normal distribution the optimized parameters obtained using a PH (water pressure head) dataset and a PH+wc dataset are located closer to the initial parameter (initial guess) with respect to the optimized parameters obtained from a water content dataset. The reason stays in the higher performance of the model when simulating a pwp distribution and the relative difficulty in obtaining an optimization using inverse analysis. The 9 indirect simulations that have overcome the third

calibration phase used a water content dataset in the inverse analysis and, as previously mentioned, the wider margins of improvement are in the simulations of the water content distribution. As already said, the optimized parameters that use a mixed dataset are close to the parameters obtained from a PH dataset; this means that the inverse analysis is driven mostly by the pressure head dataset and less by the water content dataset when a mixed dataset is used. The explanation could be found in the higher responsivity of the sensors monitoring water potential during the simulation period with respect to the sensors monitoring water content, that assures greater variations in their trend due to external hydraulic stimuli. This factor has an important role in the quality of the calibration output. From the graphical representation of the optimized parameters on a normal distribution, it has been observed that the increase or decrease of the parameter value with respect to the initial parameter depends on the type of observed dataset used: for example the hydraulic parameters θs and ks of the embankment layer show an increment with respect to the initial guess when a PH+wc dataset is used, a decrease for a wc dataset, while θs shows an increment and ks a decrement for a PH dataset. From all this, we understand the dominant influence of the observed dataset used in the objective function on the results of the inverse analysis.

The Levenberg- Marquardt algorithm depends on the set of initial parameters (initial guess) which is the starting point of the inverse analysis. In the calibration phase, all the performed indirect simulations start from a set of hydraulic parameters obtained from a campaign of laboratory tests. In order to investigate how strong is the dependence between inverse analysis output-initial guess, a campaign of indirect simulations optimizing the embankment layer using the VGM has been performed starting from a new set of initial parameters. If we suppose that laboratory tests are not available, the pedotransfer function could help to choose the initial parameters. Inverse analysis have been performed optimizing different combinations of hydraulic parameters and using different observed datasets (PH, wc, PH+wc) as done in the calibration process. The number of successful simulations is so low that no conclusions could be drawn. A general difficulty in the convergence of the indirect simulations has been noted starting from an initial set of parameters far from the ‘’good’’ one. Working on the time steps, iteration criteria and on the mesh it is possible to obtain convergence in the direct problem using a set of initial parameters of ‘’first attempt’’. The convergence of the direct simulation is necessary but not sufficient to obtain satisfactory results in the indirect simulation. It has been observed that a ‘’smooth’’ convergence of the direct simulation (low number of iteration for each time step and a short calculation time) is necessary to have a ‘’smooth’’ convergence of the inverse analysis. If the convergence of the direct problem is difficult, it means that in the indirect problem as the objective function moves away from the set of initial parameters, more the convergence is difficult to obtain. When the inverse analysis starts from a good set of initial parameters few iterations and a short calculation time is sufficient to adjust the set of parameters to the optimal one. A second attempt has been made changing only the initial values of the parameters that are optimized in each indirect simulation instead of the whole set of initial parameters. The 25% of the simulations reaches convergence, and also in this case, the majority of them use a PH dataset but the greater changes in the simulation performance have been obtained for a wc and PH+wc datasets. The three phases of performance evaluation used in the calibration phase are applied again to this new set of simulations. It has been observed that the majority of the indirect simulations remains in their local minima, while few of them approaches the minimum in which the initial set of parameters from laboratory tests is located. Due to the fact that the performance of these simulations is poorer compared to the base simulation (from laboratory tests) we can conclude that the minima in which these new optimized parameters are located are local and not global. The choice of the initial set of parameters has a strong influence on the results of

the inverse analysis and it is always suggested to repeat the calibration starting from different points of the parameters space. For more information see chapter §7_6.

In the performed calibration programme two weighting distributions have been tested: the Triangular Moving Average Technique and all the observed data with the same weight equal to unity (the simplest). On a restricted number of indirect simulations that optimize the embankment layer using the VGM, different weighting distributions have been tested. 6 different types of weighting distributions have been elaborated based on the different temporal periods (flood peak periods, off-peak periods; drawdown). The weights of the observed data in the off-peak periods are reduced with respect to the flood peak periods in different proportions because the calibration process focuses on a reliable simulation of the flooding events. In another casuistry, the weights attributed to the flood drawdown are increased with respect to the period in which the water level reaches the peak because the minor FOS are often registered with a certain delay from the peak. 3 different types of weighting distributions have been elaborated depending on the location of the sensors in the riverbank section, on the accuracy of the sensors and on the type of sensor (monitoring water potential or water content). For what concerns the position in the bank section the greatest weight is given to the sensors in the unsaturated zone while a reduced weight for the sensors below the water table and in the first meters from the soil surface because affected by the evapotranspiration contribution and less by the hydraulic parameters. For what concerns the different weights based on the type of sensor: the sensors monitoring water potential have a greater weight compared to the sensors monitoring soil water content due to their major relevance for the purposes of the work. The distribution that differentiates the weights based on the accuracy of the sensors gives more relevance to the sensors monitoring water potential especially to T8 (greater accuracy with respect to MSP6), less for SM150T and GS3 which have the same accuracy but they monitor soil water content. The same evaluation procedure in three phases used in the calibration process is replicated again. It has been observed that only the 33% of the simulations using a weighting distribution based on position, type and accuracy of the sensors reached convergence, the 37% for the weighting distributions based on the different temporal periods while more than 60% for the indirect simulations using a weighting distribution equal to 1. These results are not surprising because inverse analysis encounters difficulties in the minimization of the objective function in the most relevant instants of the simulation when great changes in pwp and water content distribution are registered in a short time. The sensors in the unsaturated zone of the section are the most responsive to external stimuli showing rapid changes during flooding events and the calibration of the hydraulic parameters must focus on these responses. It has been observed that the greater scores (and the greater enhancements of the simulation performance) belong to the simulations that use a weighting distribution based on the different temporal periods, poorer performances for the weighting distribution based on position, accuracy and type of measurement and for the group of indirect simulations that gives more weight to the flood drawdown periods. For more information see chapter §7_7.

The sets of optimized parameters that overcome the third phase of calibration are tested on a new monitoring period (24th June 2018-30th June 2019) using a new set of observed dataset (validation phase). The same procedure in three phases used for the calibration phase is applied again in the validation phase in order to test the predictive capacity of the models. In total, the performances of 10 indirect simulations are investigated. The indirect simulations show a good performance in the flood peak periods and a poor performance in the off-peak periods; so globally the performance of the model in the 2019 year is acceptable being halfway between the two trends. The third phase of the performance evaluation, shows

that the enhancements are localized in certain layers or limited to certain sensors and simulation periods, as observed also in the calibration phase. For this reason the choice of the best set/s of observation points has to be made based on the purposes of the modelling and not only on the score achieved in the evaluation process. In the presented case study, all the sets of metrics that overcome the three phases of the validation process (7/10) optimize the parameters of the embankment layer and the simulation of its hydraulic and retention behaviour is the focus of the entire modelling so the maximum score has been used as unique parameter of choice of the final sets of parameters. It must be emphasize that the model output seems to depend greatly on the initial conditions that are imported from the last temporal instant of the previous model (2018 year) and a certain gap between observed and simulated pwp distribution is observable at time zero. This influence on the trend of the observation points seems to not decrease in time due to the fact that the initial part of the 2019 simulation period does not present any flooding events able to erase or partly erase the influence of the chosen initial conditions. Hydrus2D allows to set initial conditions in a simplified manner, while importing the observed values of pwp in the measurement points and interpolating in the remaining nodes could give better results. For more information see chapter §8.

The scalar field of the Local Factor of Safety (LFS) has been used to test the incidence of the use of the set of optimized parameters obtained at the end of the calibration and validation phase instead of the set of initial parameters for simplified stability considerations. The LFS is defined as the current state of stress in each point divided for the shear strength obtained shifting the Mohr-Coulomb circle leftward until it touches the MC envelope. It is a scalar quantity computed in each point of the domain and its physical meaning is ‘’the distance’’ of the current punctual state of stress from the failure one. It has been observed, as could be expected, that the major differences in the LFS distribution between the optimized set of parameters and the initial one could be found in the embankment layer and in the berm with variation of 3-5% between the soil surface and 10m of depth, up to 22% between 10-20 m. The variations are reduced as the differences in the set of optimized parameters with respect to the initial one.

Concluding, problems of parameters correlation, convergence issues, non-identifiability are relevant aspects to take into consideration when the Levenberg-Marquardt algorithm is applied to complex models as in the considered case study. A good set of initial parameters from laboratory or in situ tests and a reliable dataset of observation points of different typology (water content, water pressure head, cumulative flux etc..) are fundamental data to obtain a good performance of the indirect simulations.

The more the model is complex and multi-parametrical, the more the modeller must focus only on the relevant aspects of the analysis keeping in mind the main goals of the simulation he is performing: localized enhancements of the model performance (focused on few layers and few observation points), optimization of the most sensitive and relevant parameters of the simulation are necessary simplifications to obtain good results in the indirect approach.

If the calibration concerns simple parametric models, inverse analysis based on the L-M algorithm is a valid optimization technique, as furtherly discussed with the RMS project.

If the calibration concerns complex multi-parametric models or multi-layered domains, as in the present case study, the use of inverse analysis could be demanding due to the high number of inverse analysis to perform. Moreover often the indirect approach leads to enhancements in the model performance of small entity, localized or which affects only a restricted number of layers.

For all these reasons, optimization techniques are moving toward **Evolutionary Algorithms** (E.A) which differ from traditional algorithm because they are dynamic and not static (they can evolve over time). Among the E.A, the Genetic Algorithm (G.A) based on the Darwin's theory of evolution (natural selection and genetics) assures a quicker and easier convergence of multi-parametrical models to the best solution.

PART II

1_PROJECT BACKGROUND AND AIMS

This part of the Ph.D project focuses on the stabilizing effect of plant roots on soil slopes and it originates from a real stability problem that occurred on a site close to the University of Newcastle, NSW, Australia. A partially back-filled slope behind a retaining wall of the Newcastle Inner City Bypass Cut 7, West Charlestown (Newcastle, Australia) experienced shallow instability in March 2017. The back slope has a 1:1 batter (45%) and consists of an A-Horizon (topsoil) and B-Horizon (subsoil). The instability is believed to have been triggered by heavy rainfalls, despite the presence of drains above and within the slope. Figure 43 shows an aerial view of site (noted A), a view of the slope and the retaining wall (noted C, D, E) and a view of the slope with scars due to the recent instabilities (noted B). An in situ inspection conducted in November 2019 showed a new instability which led to the accumulation of material at the toe of the slope (just above the crest of the retaining wall). Aerial images taken from October 2007 to November 2018 show that the tree population has decreased over time, reducing the degree of reinforcement of the soil by roots (Figure 44 from Report 11/12/2018, SESL Australia).

The Road and Maritime Services (RMS) of New South Wales, governmental agency responsible for building and maintaining road infrastructure and safety for roads and waterways, suggested the use of long stem planting using native plants (Melaleuca Styphelioides) to a depth of 1m to stabilize the slope. The stabilization effect is expected to be twofold: the stem and roots will act as mechanical reinforcement and the water uptake by the roots will increase suction in the soil, resulting in higher soil strength. This latter effect is the focus of the present study.

The project main purpose is to investigate the effects of plant evapo-transpiration on the spatial and temporal distribution of pore pressure (suction) by means of a large-scale evapo-transpiration apparatus. Changes in suction induced by a plant at certain depth intervals and radial distance from the root bulb were measured in laboratory and reproduced numerically using the commercial code Hydrus 2D.

The significance of root water uptake for vadose processes has long been recognised (Jarvis, 1989; Wu et al, 1999; Javaux et al, 2008; Javaux et al, 2013). Adequately modelling root water uptake is required to accurately estimate moisture movement in the ground and to the atmosphere and assess the effect that living plants have on surrounding soil (Vrugt et al, 2001; Canadell et al, 1996; Somma et al, 1998). It is also relevant if one wishes to accounts for the stabilizing effect plants can have on slope stability.

Give a slope geometry and plant species, it is possible to simulate and validate a 2D root water uptake model and incorporate it into a numerical multidimensional flow and stability model. Such approach allows the simulation of complex transient environmental phenomena i.e. river banks subjected to transient river water flux and atmospheric conditions, with vegetated slopes etc.

The calibration of a RWU numerical model simulating the tests performed with the large-scale apparatus is a first step towards quantifying the possible stabilization effects of Melaleuca Styphelioides for the West Charlestown bypass.

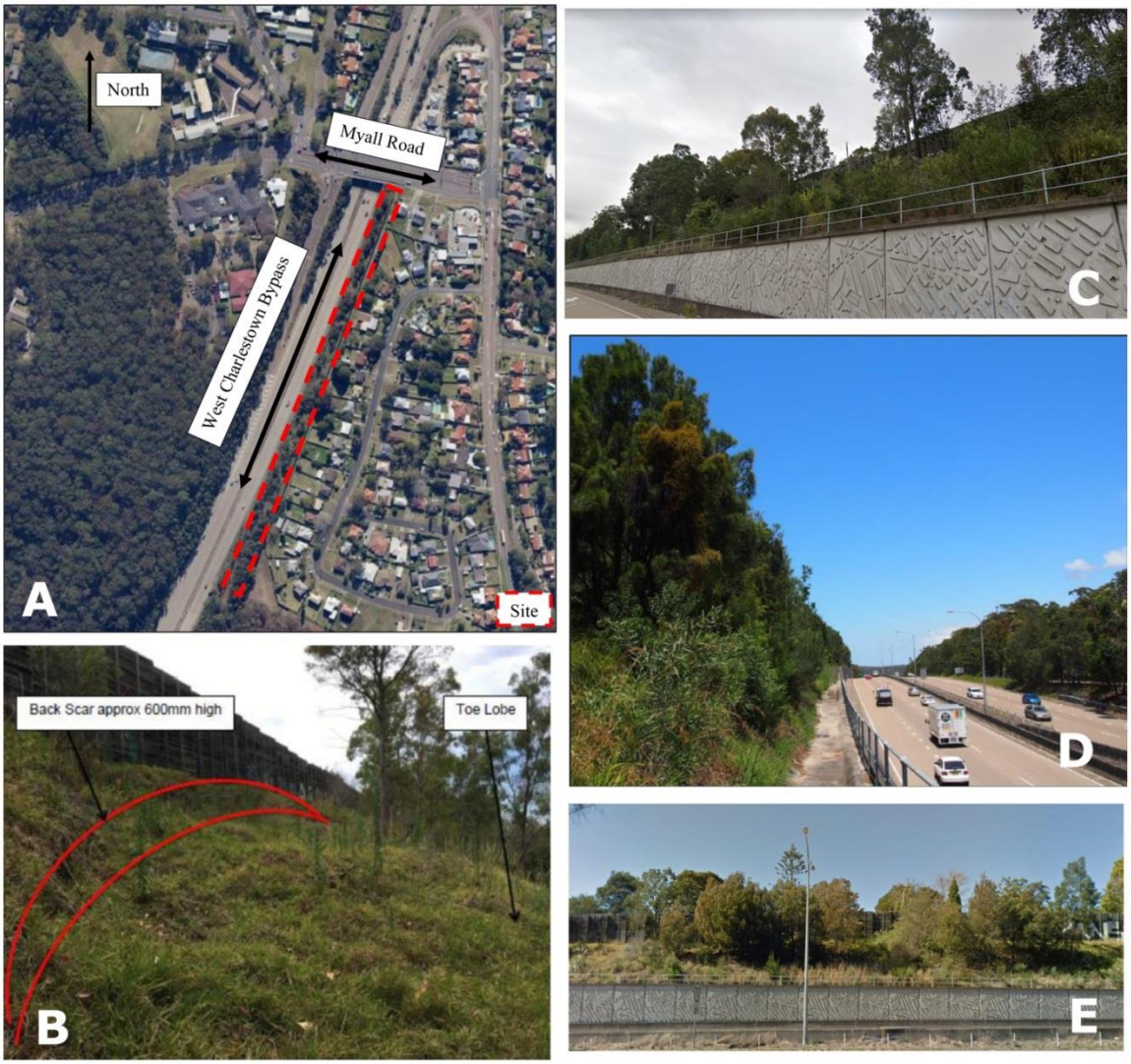


Figure 43: Aerial view of the site (A), the slope and the retaining wall (C,D,E) and cracks due to shallow instability (B).

Figure 44 Aerial images from October 2007 to November 2018 of the investigated slope (The Newcastle Inner City Bypass Cut 7, West Charlestown, Newcastle, Australia) from Report 11/12/2018, SESL Australia.

Measuring soil moisture around root bulbs has been done in the past (see literature review) but the experimental work conducted in the present thesis still present some innovations:

- The size of the laboratory experiment conducted is a half-way scale between typical small-scale tests and large-scale field tests. It offers controlled conditions while eliminating edge effects. As a consequence, modelling the evapo-transpiration phenomenon is more accurate and still representative of real conditions.

- Water was provided from below as groundwater as opposed to being supplied from the top, as rainfall or irrigation, which is commonly done in the literature. This allows to simulate the change in the water level due to rainfall events as it happens commonly in nature in non-cultivated/non-irrigated ill-slopes as riverbanks.

- Negative pore water pressure was measured, as opposed to moisture content, to calibrate the RWU spatial distribution of the numerical model. No similar case studies using this type of observed information are to the knowledge of the authors. As reported in chapter §2_1_4 suction measurements in close proximity of root bulbs have been performed since the eighties using tensiometers but such experimental data have not been used to calibrate any hydraulic model which takes into consideration the RWU distribution of the investigated plants. Differently volumetric water content information from TDR sensors have been used in model calibration (for more information see chapter §2_1_3).

- No such model was ever calibrated for the Melaleuca Styphelioides, a native plant of New South Wales, Australia (see Figure 3).

Figure 45: The adult stage of the Malaleuca Stypelioides native plant (left side), a photo of the young plant used for the laboratory experiments (centre) and the typical white flowers of the adult plant (right side)

2_LITERATURE REVIEW

Soil bioengineering is becoming increasingly used and acknowledged in riverbank restoration and in the management of upland slopes. Plants are used worldwide to stabilize slopes or prevent erosion in different climatic zones (Morgan and Rickson, 2003; Wu et al, 2015). The natural bio-engineering techniques are globally appreciated for the relative low cost, aesthetic, sustainability and because they are environment-friendly solutions. Soil bioengineering is usually achieved by hydro seeding (application of an emulsion of seeds, fertilizer, wood fibres) or direct pit planting of young trees. The former technique is used to control erosion by water and wind (Vaciago, 2013) while the latter helps improving slope stability. Indeed, slope vegetation is proved to be an efficient technique to mitigate erosive phenomena and increase slope stability (Wu et al, 1979, Mickovski et al 2009). The type of vegetation to be used in soil bio-engineering has to be carefully selected. Grass cover is used to reduce the coefficient of erodibility of the soil and increase slope durability (Zhu et al, 2017) while trees are more effective in water content decrease due to a higher transpiration activity (Garg et al, 2015). It is always important to consider the environmental context in which plants will grow: native vegetation should always be preferred because it ensures a higher probability of survival in the long period (Schiechtl and Stern, 1996).

Rainfall-induced landslides are a global hazard costing society in terms of human lives and economic losses (Sidle and Bogaard, 2016). The likelihood of rainfall-induced landslides is expected to increase because of climate change and the need for adequate mitigation and remediation measures is fast growing.

Vegetation enhances soil shear strength through mechanical and hydrological effects which are going to be discussed in this chapter. The latter effect, although acknowledged, lacks a reliable and rigorous quantification in the scientific community (Stokes et al, 2014).

Plant roots can possess high tensile strength and act as soil reinforcement, providing an apparent cohesion to the soil. Moreover, strong roots could grow across failure surface, working as anchoring points. This type of contribution depends on root architecture, root mass and can significantly change with time (Wu et al, 2015). A relevant degree of uncertainty is present because it is difficult to know the lateral spread, depth and architecture of a root system within a soil mass in practical applications.

Vegetation interacts in different way with the atmosphere and the medium in which plants live. During a wetting event (rainfall), vegetation intercepts some precipitation thus reducing the amount of water infiltrating the soil. However, it is possible for water to flow along the stem and rapidly infiltrate the soil through root channels leading in case of a vegetated slope to piping failure (Liang et al, 2011). During drying (following rainfall), transpiration generates a water flow from the soil to the environment that reduces the degree of saturation and increases soil suction. Since suction is one of the two independent stress variables governing the mechanical behaviour of unsaturated soils (Burland, 1965), it is very relevant to take into consideration transpiration effects to properly predict the soil response.

Another phenomenon attributed to roots is the modification of the pore space geometry of soil, for example roots could block pore space or create macro-pores due to root's decay, leading to important changes in the hydraulic permeability of the soil (Scanlan and Hinz, 2010). Vegetation as grass carpet is often responsible of shallow landslides because below root tips the permeability is lower than in the soil permeated by roots. Because of drying and shrinkage, it is also possible for roots to loosen the subsoil opening flow channels that drain the water from the root zone (Liang et al, 2011).

Vegetation also indirectly affects water infiltration in the soil. First, the presence of plant canopy reduces the surface runoff and the water infiltration. Vegetation is able to reduce the impact energy of rain drops preventing splash erosion and to slow down the surface runoff. This beneficial effect of plant roots in erosion control is worldwide acknowledge but difficult to quantify because it depends on climatic conditions, vegetation cover, species and soil conditions (Wu et al, 2011). Finally, vegetation can affect the amount of evaporation through the soil surface because of falling leaves cover the soil.

A detailed study of evapotranspiration effects on slope stability is possible by coupling the mechanical system (soil+plant) to the hydrological system, accounting for the atmospheric boundary conditions (Wu et al, 2015). Considering the phenomena of water infiltration, runoff, evapotranspiration and root water uptake in a model is extremely complex and requires knowledge in the fields of hydraulics, soil mechanics, agronomy and soil physics.

2_1_INFLUENCE OF EVAPOTRANSPIRATION ON SUCTION DISTRIBUTION

Evapotranspiration is defined as the exchange of vapour water between a vegetated surface and the atmosphere. The potential evapotranspiration is the maximum value that could be reached when no restriction to water vapour exchange is present. In order to pass from the potential transpiration to the actual transpiration, plant characteristics, soil moisture conditions and atmospheric conditions have to be considered.

A key phenomenon limiting evapotranspiration is the ability of the plant to suck water from the soil, which is known as root water uptake (RWU). RWU is the process by which root cells extract water from the soil by osmosis (due to their low solute potential) and by capillary action water penetrates the xylem (vascular tissue in the internal part of the root) which is able to transport water and the nutrients dissolved in it to stem and leaves (shoot). Water is driven from the roots xylem to the leaves (stomata) by capillary action and primarily by differences in water potential. The quantitative estimation of RWU contribution is of primary importance for an efficient management of water and nutrient resources and to reliably estimate crop water need (Cai et al, 2017; Molz, 1981). However, the computation of RWU is challenging due to uncertainties around the properties of the root system and of water uptake mechanisms.

2_1_1 Root water uptake models

The Penman-Monteith equation (introduced in chapter §5, PART I) is commonly used to compute the magnitude of evapotranspiration. Soil suction is a parameter that significantly influences the rate of water uptake by roots (Cardon et al, 1992; Zuo et al, 2006; Leung et al, 2015; Ni et al, 2020): for suctions greater than 1500 kPa (known as ‘’the permanent wilting point’), plants cannot extract water from the soil. In contrast, after a wetting process and after excess water has drained away and the moisture movement towards deeper strata has ceased (a point known as field capacity), water extraction from the plant is high. Field capacity sits around 10 to 30 kPa of suction.

The first models describing water uptake by roots are based on Ohm’s law, a 1D steady state flow in which the transpiration rate T is expressed in terms of water potential Ψ and the hydraulic resistance of the soil and the plant.

$$T = \Psi_S - \Psi_R R_s = \Psi_R - \Psi_R R_p \qquad \text{[equation 1]}$$

where Ψ_S, Ψ_R are the water potentials in the soil and at the root surface respectively and R_s and R_p are the resistance to the flow in the soil and in the plant. Because Ψ_R is extremely difficult to obtain, the equation is often re-written in terms of Ψ_l , the water potential in the plant leaves[4].

$$T = \Psi_S - \Psi_L R_s + R_p \qquad \text{[equation 2]}$$

One difficulty of equation 2 lies with the determination of the hydraulic resistance parameters R_p and R_s (Wei Wu et al, 2015; Glinski and Lipiec, 1990). Gardner (1960) proposed the single root model or microscopic model, an approach inspired by Darcy's law and applied to root water uptake. In this approach, a single root is simplified as an uniform infinitely long cylinder of radius r that extracts water at a steady state flow at a maximum distance c from its centre (c is half of the distance to another root considered as a cylinder too). Half-mean distance is used as an approximation when real spatial information are missing in many modelling approaches and could be estimated from the root length density L_z.Water is radially absorbed along the root and the following water extraction function prevails:

$$\Psi_s - \Psi_R = \frac{q_r}{4\pi k} \ln\left(\frac{c^2}{r^2}\right) \qquad \text{[equation 3]}$$

$$c = (\pi\, L_z)^{-1/2} \qquad \text{[equation 4]}$$

where Ψ_S, Ψ_R are the water potentials in the soil and at the root surface respectively (MPa), L_z is a root length density (m of root/ m^3 of soil), q_r is the water extracted from the soil (m^3 of water/m of root per sec), k is the hydraulic conductivity of the unsaturated soil ($m^2 s^{-1} MPa^{-1}$). The Gardner (1960) approach has been criticized because the required measurements are impossible to be performed around an individual root with the nowadays techniques.

More recently, due to the difficulty of extending the microscopic approach of Gardner (1960) to the complexity of a real root system, a ''macroscopic'' function to describe the root uptake has been formulated. While all the microscopic models elaborated from Gardner's intuition investigate the microscopic structure i.e. the individual roots, the macroscopic approach considers the water uptake from the root zone as a whole. This approach consists on incorporating in the Darcy-Richards equation, that describes flow in soils, a ''sink term'' $S(z)$ that expresses the volume of water absorbed by roots per unit volume of bulk soil in time unit ($m^3/(m^3$ *sec)). $S(z)$ in the macroscopic model is the analogous of q_r ($m^3/(m^3$ sec) in Gardner microscopic model. The sink term $S(z)$ integrated over the rooting depth gives the actual transpiration in length per time (m/sec). The water flow in the soil interested by a RWU phenomenon can be computed by adding the sink term $S(z)$ to the Richards' equation (Cardon and Letey, 1992) as follows (2D formulation):

$$\frac{\partial}{\partial x}\left(k\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial z}\left(k\frac{\partial h}{\partial z}\right) = \frac{\partial \theta}{\partial t} + S(z) \qquad \text{[equation 5]}$$

Where h is the pressure head (m), x and z are the coordinated in the 2D space, k is the unsaturated hydraulic permeability, θ is the volumetric water content and $S(z)$ is the sink term.

[4] Leaf water potential (LWP) allows to measure the plant water status during the day in reaction to a change in the root water absorption and leaf transpiration. LWP is obtained by means of a pressure chamber (Scholander et al, 1965) based on the pressurization of a leaf with a neutral gas in order to estimate the capacity of the leaf cells to retain water.

Based on how the sink term S is handled, models fall in one of the two following groups:

2_1_1_1 Type I model

The type I model describes the physical phenomenon as a water flow from the soil to the roots and it moves from the work of Gardner (1964). The Type I model relies on gradient type expressions which take the form of:

$$S = BKG \qquad \text{[equation 6]}$$

where K is the conductivity term (in m/day), G is the water gradient from the soil to the root expressed as the difference between root and soil water pressure head (in m) and B is a geometrical term (m^{-2}). Starting from the work of Gardner (1964), other type I models were elaborated as Whisler et al (1968); Feddes et al (1974); Hillel et al (1976), among others. The major difficulty in the application of Type I model stays in the experimental difficulties involved in the determination of the empirical root function B that pushed the research toward the elaboration of Type II models (Feddes et al, 1976).

2_1_1_2 Type II model

The type II model describes the physical phenomenon using empirical functions and it considers RWU as a response to water potential. The expression of the sink term S has the form of:

$$S = \alpha(h)S_{max} \qquad \text{[equation 7]}$$

where $\alpha(h)$ is a dimensionless stress response function equivalent to the ratio between actual and maximum root uptake (S_{max}) and it ranges between 0 and 1. h indicates the dependence of α to the water pressure head. Several models use this type of approach to the sink term S as Feddes et al (1978); Hoogland et al (1981); van Genutchen (1987), among others.

Feddes et al (1978) assumes the **maximum sink term (S_{max})** as the potential transpiration over the depth of the roots and it is expressed as follows:

$$S_{max} = \frac{T_p}{|z_R|} \qquad \text{[equation 8]}$$

Where $\boldsymbol{T_p}$ **is the potential transpiration rate** (m sec^{-1}) and $|z_R|$ is the root-zone depth (m).

Hoogland et al (1981) assumes S_{max} as a function of the soil depth, in which case, the sink term S at the given depth z is expressed as:

$$S(z) = T_{p\,max}\beta(z)\alpha(h) \qquad \text{[equation 9]}$$

where $\boldsymbol{T_{p\,max}}$ **(m/sec^{-1}) is the maximum potential transpiration rate**, β(z) (1/m) is a shape parameter function of the root distribution at the depth z and $\alpha(h)$ is a function assuming value in the range [0,1] to express the transpiration reduction due to the value of soil water pressure head h. Over the last decades, many shape functions $\beta(z)$ have been proposed (uniformly, triangular, parabolically distributed roots), all of them are idealised root models based on the distribution of the root biomass. The **actual transpiration rate** $\boldsymbol{T_a}$ is obtained integrating $S(z)$ from the surface to the maximum rooting depth (z_R), as follows:

$$T_a = \int_0^{z_R} S(z)dz \qquad \text{[equation 10]}$$

Type II models suffer from data uncertainty and they are rarely based on plant physiology. As a consequence, they possess low predictive capability; yet they are widely used (Zhu et al, 2015).

2_1_2 Root water uptake in Hydrus 2D

Hydrus 2D by Pc-Progress has been presented in chapter §2_7 (PART I), with general information and the governing equations to simulate water movement in partially saturated media. In this chapter, attention is focused on the root water uptake formulation in the code. Hydrus uses the macroscopic approach to compute RWU where the potential transpiration is distributed proportionally to the root density in the root zone. The basic assumption is that the amount of water extracted from the soil depends on the amount of roots present in terms of mass or surface area.

2_1_2_1 Computation of the Potential Transpiration

Potential transpiration T_p, as stated before, depends only on environmental conditions and it can be computed using the Penman-Monteith equation. The potential transpiration T_p [LT^{-1}] is distributed on a 2D root domain according to the following expression:

$$S_p(x,z,t) = \frac{1}{L_x L_z} L_t T_p(t) \qquad \text{[equation 11]}$$

Where S_p [T^{-1}] is the potential root water uptake in the point (x,z) of the domain and at the time t. L_x [L] and L_z [L] are the depth and width of the root zone, respectively. L_t [L] is the width of the soil surface associated with the plant transpiration [L].

The potential transpiration T_p reduces with suction and salinity of the soil to reflect the assumption that water absorption by roots decreases as the soil becomes dryer and salinity concentration increases. In this approach, the sink term is inserted in the governing equations of the mass balance (Simunek et al, 2008). Note that this approach neglects the effect due to root geometry and the possible flow pathways around the roots.

In literature, a number of root water uptake reduction function α(h, h_ϕ, x,y,z,t) have been proposed: from the S-shape function presented by van Genutchen (1987) that requires three parameters, to more complex formulations such as the one by Feddes et al (1978) requiring five parameters. The root water uptake reduction function α depends on the considered spatial point of the domain (x,z in the 2D space);on the temporal instant (t); on the soil water pressure head (h) and osmotic pressure head (h_ϕ).

In cases where the distribution of RWU over the root zone is not uniform and the root zone has an arbitrary shape, the expression of the potential transpiration (equation 11) can be re-written as:

$$S_p(x,z,t) = b(x,z,t) L_t T_p(t) \qquad \text{[equation 12]}$$

where $b(x,z,t)$ is the normalized water uptake distribution [L^{-2} in the 2D problem],that is function of space and time in order to take into consideration the plant root growth in space and time. $b(x,z,t)$ can be linear (Feddes et al, 1978), exponential (Raats, 1974) or more flexible as in Vrugt et al (2001) because it allows spatial variations of water uptake due to non-uniform water application and root length density patterns. The function $b(x,z,t)$ has to be normalised to ensure that its integral over the root zone is equal to one as in equation 13.

$$\int_{\Omega_R} b(x,z,t)d\Omega = 1$$ [equation 13]

The normalisation of the function $b(x,z,t)$ is performed as follows:

$$b(x,z,t) = \frac{b\prime(x,z,t)}{\int_{\Omega_R} b'(x,z,t)d\Omega}$$ [equation 14]

Where $b'(x,z,t)$ is an arbitrary prescribed distribution function and Ω_R is the region occupied by the root zone [L^2].

For a 2D problem, the potential transpiration reads:

$$T_p(t) = \frac{1}{L_t}\int_{\Omega_R} s_p(x,z,t)d\Omega$$ [equation 15]

Where Ω_R is the root zone [L^2].

2_1_2_2 The root water uptake spatial distribution function of Vrugt et al (2001)

A discussion of the 2D root water uptake spatial distribution function of Vrugt et al (2001) implemented in Hydrus 2D will be presented briefly here. Vrugt's function is based on the exponential function by Raats (1974) (see equation 16) and could be formulated as in equation 17.

$$b(x,z) = (1 - \frac{z}{Z_m})e^{-\frac{p_z}{z_m}|z^*-z|}$$ with z ≥ 0 [equation 16]

$$b(x,z) = (1 - \frac{z}{Z_m})(1 - \frac{x}{X_m})e^{(\frac{p_z}{Z_m}|z^*-z|+\frac{p_r}{X_m}|x^*-x|)}$$ [equation 17]

Raats (1974) model characterizes the dimensionless spatial distribution function $b(x,z)$ in the vertical direction (z) and Vrugt et al (2001) extends the model in the radial direction (x). The use of (x,z) in equations 16 and 17 denotes that $b(x,z)$ is defined in each point (x,z) of the RWU domain. Z_m is the maximum rooting depth [L], X_m is the maximum rooting length [L] in the horizontal direction; p_z and p_r are empirical parameters (-); z^* and $x^*$are other empirical parameters [L]. Parameters p_z and p_r are set to unity for values of $z > z^*$ and $x > x^*$, respectively. Globally the Vrugt's model requires the definition of six parameters $(Z_m, X_m, p_z, p_r, z^*, x^*)$.

2_1_2_3 The compensated root water uptake

Hydrus accounts for a **compensation mechanism** to estimate the actual transpiration rate ($T_{p\ actual}$). The compensation mechanism is supported by experimental evidence (Leib et al, 2006) according to which water uptake is increased in a less-stressed part of the root structure, if water uptake is reduced in another part of the root zone (for example because of salinity or higher soil suction).

Hydrus manages the compensation mechanism via the dimensionless water stress index (ω) defined as the ratio of actual transpiration over potential transpiration (Jarvis, 1989) and the critical water stress index (ω_c), defined as the threshold above which water uptake stops in the more-stressed zone and the less-stressed parts compensate fully (actual transpiration is equal to potential transpiration).

For $\omega>\omega_c$ the **actual compensated transpiration rate** T_{ac} is defined as the ratio of the **actual transpiration rate** $\boldsymbol{T_a}$ to the dimensionless water stress index ω as follows:

$$T_{ac}(t) = \frac{T_a(t)}{\omega} \quad \text{[equation 18]}$$

$$\frac{T_{ac}(t)}{T_p(t)} = \frac{T_a}{T_p(t)\omega(t)} = \frac{\int_{\Omega_R} \alpha(h,x,z,t)b(x,z,t)d\Omega}{\omega(t)} = \frac{\omega(t)}{\omega(t)} = 1 \quad \text{[equation 19]}$$

where $b(x,z,t)$ is the spatial root distribution function, $\alpha(h,x,z,t)$ the water stress response function, $\omega(t)$ the water stress index, Ω_R is the root zone [L^2], $T_p(t)$ the potential transpiration function.

In this case, following equation 12, **the compensated root water uptake** $s_c(h,x,z,t)$ [T^{-1}] is defined as:

$$s_c(h,x,z,t) = \alpha(h,x,z,t)b(x,z,t)L_t\frac{T_p(t)}{\omega(t)} \quad \text{[equation 20]}$$

where L_t [L] is the width of the soil surface associated with the plant transpiration [L], $b(x,z,t)$ is the spatial root distribution function, $\alpha(h,x,z,t)$ the water stress response function, $\omega(t)$ the water stress index.

For $\omega<\omega_c$, the **actual compensated transpiration rate** T_{ac} is defined as the ratio of the actual transpiration $T_{p\ actual}$ to the dimensionless critical water stress index ω_c as follows:

$$\frac{T_{ac}(t)}{T_p(t)} = \frac{T_a}{T_p(t)\omega_c} = \frac{\int_{\Omega_R} \alpha(h,x,z,t)b(x,z,t)d\Omega}{\omega_c} = \frac{\omega(t)}{\omega_c} < 1 \quad \text{[equation 21]}$$

where $b(x,z,t)$ is the spatial root distribution function, $\alpha(h,x,z,t)$ the water stress response function, ω_c the critical water stress index, Ω_R is the root zone [L^2], $T_p(t)$ the potential transpiration function.

In this case, following equation 12, the **compensated root water uptake** $s_c(h,x,z,t)$ [T^{-1}] is defined as:

$$s_c(h,x,z,t) = \alpha(h,x,z,t)b(x,z,t)L_t\frac{T_p(t)}{\omega_c} \quad \text{[equation 22]}$$

The compensation is maximum in the part of the root structure where the water uptake is optimal and equal to zero above the anaerobiosis point and below the wilting point, in the other points it is proportional to the water stress. For agricultural plants, ω_c is a very high value and the compensation ability is very low. In contrast, plants that live in arid zones have a high ability to compensate (low ω_c) (Simunek, 2009).

Figure 46 shows the evolution of the ratio of actual transpiration rate $T_{ac}(t)$ over the potential transpiration rate $T_p(t)$ as function of the water stress index $\omega(t)$ for compensated and non-compensated models. The left axis refers to the compensated uptake, the right axis to the uncompensated one (follow the direction of the arrows). The reduction of actual transpiration rate in the compensated model is smaller than in the uncompensated model.

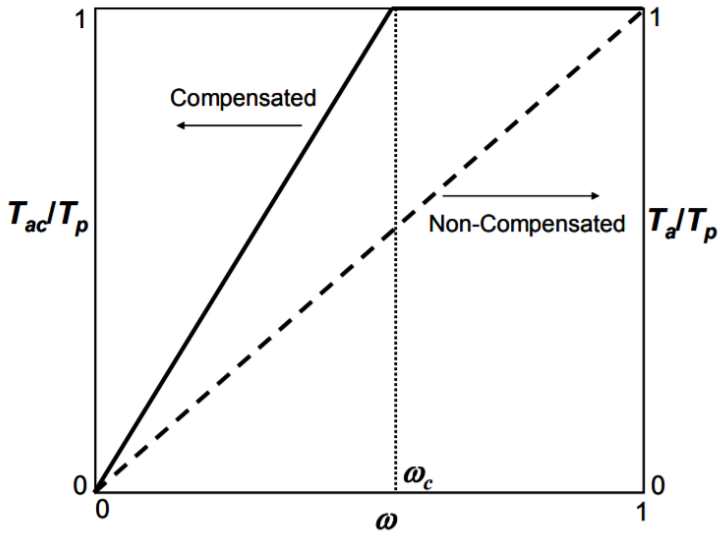


Figure 46 The dimensionless water stress index $\omega(t)$ in function of the ratio of actual compensated transpiration rate ($T_{ac}(t)$) over potential transpiration rate ($T_p(t)$) for compensated model (left axis) while the actual transpiration rate ($T_a(t)$) over potential transpiration rate ($T_p(t)$) for non-compensated model (right axis). From Simunek et al, 2009.

The analytical formulation of Figure 46 is the following:

$$\frac{T_{ac}(t)}{T_{pot}} = 1 \quad for\ \omega_c < \omega \leq 1 \qquad \text{[equation 23]}$$

$$\frac{T_{ac}(t)}{T_{pot}} = \frac{w}{w_c} \quad for\ \omega < \omega_c \qquad \text{[equation 24]}$$

As long as the plant stress indexω is above a threshold value ω_c the actual transpiration rate $T_{ac}(t)$ is equal to the potential transpiration rate T_{pot} while it decreases linearly when ω is below ω_c.When the compensated mechanism is not considered ω is set equal to 1.

2_1_2_4 The uncompensated root water uptake

If the compensation mechanism is not taken into consideration, the **actual root water uptake** is computed multiplying the potential root water uptake $s_p(t)$ [T^{-1}] by the stress response function $\alpha(h, h_\phi, x, z, t)$, as follows:

$$s(h, h_\phi, x, z, t) = \alpha(h, h_\phi, x, z, t) s_p(t) \qquad \text{[equation 25]}$$

$s_p(t)$ has been defined in equation 12.

The actual transpiration rate $T_a(t)$in absence of compensation can be written as follows:

$$T_a(t) = \frac{1}{L_t} \int_{\Omega_R} s(h, h_\phi, x, z, t) d\Omega = T_p(t) \int_{\Omega_R} \alpha(h, h_\phi, x, z, t) b(x, z, t) d\Omega \qquad \text{[equation 26]}$$

where $s(h, h_\phi, x, z, t)$ is the actual root water uptake[T^{-1}], $b(x, z, t)$ is the spatial root distribution function, Ω_R the region occupied by the root zone [L^2]

2_1_2_5 The stress response function of Feddes et al (1978) and van Genuchten (1987).

Assuming the stress response function of **Feddes et al (1978)**, the reduction in potential root water uptake with water pressure head has the shape of Figure 47 (a) from Simunek et al (2005) and it has the following expression:

$\alpha(h) = 0 \; for \; h \geq h_1, h \leq h_4$ [equation 27]

$\alpha(h) = \frac{h-h_1}{h_2-h_1} \; for \; h_2 \leq h \leq h_1$ [equation 28]

$\alpha(h) = 1 \; for \; h_3 < h \leq h_2$ [equation 29]

$\alpha(h) = \frac{h-h_4}{h_3-h_4} \; for \; h_4 < h \leq h_3$ [equation 30]

h_1 is defined as an arbitrary anaerobiosis point and for h>h_1 the soil is close to saturation and the transpiration phenomenon is considered to be nul (α=0) due to oxygen lack. h_4 is defined as the wilting point head and for greater values of pressure the transpiration is zero. Between h_1-h_2 and h_3-h_4 the water uptake increases/decreases linearly with h. In the interval between h_2 and h_3 the potential root water uptake is equal to the actual root water uptake while for value below h_3 (h_{3low} or h_{3high}) the uptake is reduced to the actual root water uptake and the reduction start is dependent on the rate of transpiration. In Hydrus Feddes' stress response function is implemented (Figure 47(a)) together with a database of parameters for agricultural crops.

It is relevant to stress that $\alpha(h)$ depends also on the potential transpiration rate T_p because h_3 is a piecewise linear function of T_p:

$h_3 = h_{3high} \qquad for \; T_p > T_{p_1}$ [equation 31]

$h_3 = h_{3high} + \frac{(h_{3low} - h_{3high})}{(h_{3high} - h_{3low})} \; for \; T_{p_2} < T_p < T_{p_1}$ [equation 32]

$h_3 = h_{3low} \qquad for \; T_p < T_{p_2}$ [equation 33]

Where h_3 is smaller (h_{3low}, more negative) for lower potential transpiration rate (T_{p_2}) and higher for greater potential transpiration rate (T_{p_1}). h_3 depends on T_p because when the transpiration rate is higher (T_{p_1}) stomatal closure is triggered at higher soil water potential than under lower transpiration rates (Simunek et al, 2013). An explanatory scheme could be found in Figure 48.

Hydrus 2D currently implements the same linear interpolation scheme of Feddes et al (1978) in Figure 47.

Figure 47 (b) shows the water stress response function proposed by **van Genuchten (1987)**. This S-shaped function predicts the maximum uptake under saturated conditions, unlike Feddes' model which predicts a value of 0 when the root bulb is in anaerobiosis conditions. This simplification is justified only when complete saturation occurs for a short period of time. van Genuchten's function only requires two parameters: the coefficient h_{50} corresponding to a pressure head reduction by 50% and the pressure head at wilting point. Both models are empirical and not based on plant physiology (Shah et al, 2008).

The S-shaped function of van Genuchten (1987) implemented in Hydrus 2D has the following equation:

$$\alpha(h, h_\phi) = \frac{1}{1+(\frac{h}{h_{50}})^{p_1}} \frac{1}{1+(\frac{h}{h_{\phi 50}})^{p_2}}$$ [equation 34]

The van Genuchten (1987) S-shaped function takes into consideration both the reduction due to water stress (first term on the right side in equation 34) and due to osmotic stress (second term on the right side). p_1 and p_2 are experimental constants, h_{50} is the pressure head at which water extraction is reduced by 50% in condition of negligible osmotic stress, $h_{\phi 50}$ is the osmotic head at which water extraction is reduced by 50% in condition of negligible water stress.

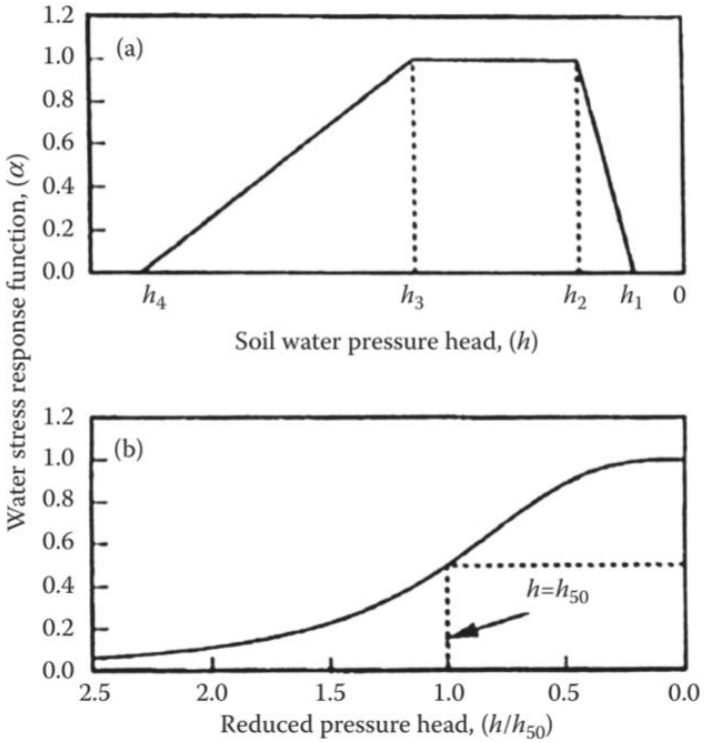


Figure 47 Water stress reduction functions α(h) by Feddes et al (1978) (a) and Van Genutchen (1987) (b) from Simunek et al (2005).

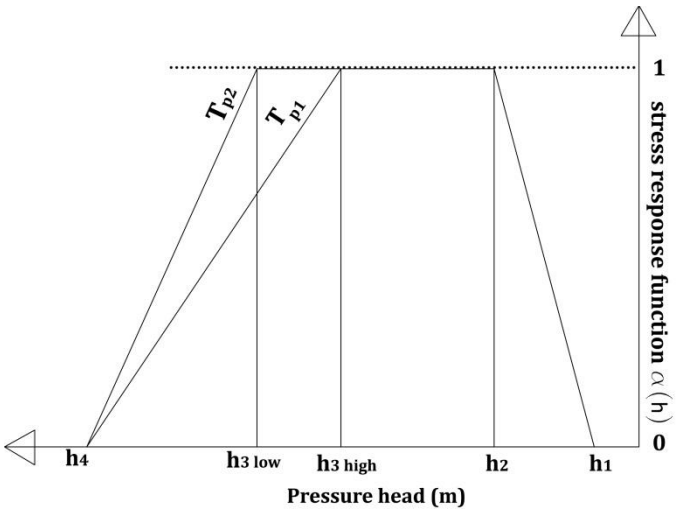


Figure 48: Graphical representation of the stress response function of Feddes et al (1978) with indication of all its parameters.

2_1_3 Experimental measurement of moisture distribution around root bulbs

The knowledge of the **root distribution** is one of the main source of uncertainties encountered in the root water uptake modelling. Traditionally rooting patterns are analysed by means of root geometrical attributes such as the **root length density** (RLD), the **root weight density** (RWD) and less commonly the **total root length** or **root diameters**. RLD is the most used because the finer roots (the most active in terms of uptake) receive higher weighting compared to large roots which are almost inactive (Coelho and Or, 1999, Zuo et al, 2004) while RWD method is almost insensitive to the active root length/area. Both the methods, RLD and RWD, are based on the in situ extraction of root samples which are then analysed with digital image cameras. It is a known fact that RLD is an important indicator of the potential RWU (Sharp and Davis, 1985). As an example, Coelho and Or (1996) proposed parametric models based on bivariate gaussian root distribution density functions (normal, semi-lognormal and lognormal) based on RLD data to describe the root uptake pattern of a drip-irrigated corn.

In situ water content measurements by means of installed sensors, result a valid alternative to the use of destructive methods (coring) to obtain information on the root water uptake spatial distribution. The use of water content information for the calibration or validation of numerical models focused on the simulation of the evapo-transpiration contribution of plants is not an innovative topic.

We can report, among other authors, **Coelho and Or (1999)** who used TDR devices installed at various depth in the investigated wetted profile cultivated with corn and they correlated the variation in water content with the local distribution of the RWU.

Green et al (2006) investigated the effect of root-water uptake and transpiration of kiwifruit vines collecting in situ data by means of TDR sensors. The changing water content pattern is able to give information not only on the timing of irrigation or rainfall events but also on the way the root bulb access water in order to meet its transpiration demand. It has been observed that kiwifruit plant is able to adjust very rapidly its absorption capacity in consequence of moisture stress (compensated RWU mechanism). For this reason the root density in the soil is an important indicator of the uptake activity but it is not an indicator of roots effectiveness when environmental stresses act.

Vrugt et al (2001) has undertaken an optimization by genetic algorithm of the root water uptake parameters using as input observation points for the objective function the water content measured around a sprinkler-irrigated almond tree.

Zuo et al (2006) correlated the RLD pattern of the winter wheat subjected to different irrigation typologies with water content data obtained by TDR sensors.

Bufon et al (2012) simulated the root water uptake phenomenon during the cotton growth under different environmental conditions and irrigation schedule and compared the obtained simulations with in situ water content measurements. 30 time domain reflectometry probes were installed for this purpose. The commercial program Hydrus 2D has been used to perform the modelling and the validation of the model. The overall simulation of the volumetric water content is in good agreement with the experimental results despite the complexity of the model: heterogeneous soil properties, multiple irrigation events, relevant evaporation contribution and the static root spatial distribution which is the most notably limit of the code due to the impossibility to give in input a growing rooting system.

Deb et al (2013) evaluated the spatial-temporal distribution of the compensated/uncompensated root water uptake of a pecan tree that could be used in a variety of different scenarios from the management of the irrigation to the understanding of the involved fluxes to the water table and to the atmosphere. The root distribution has been elaborated starting from root length density (RLD) information obtained extracting soil cores. The model calibration and validation have been performed by means of Hydrus 2D using measured data of volumetric water content obtained from time domain reflectometry (TDR, Campbell Scientific Inc.) sensors at different radial distances from the tree trunk and at different depths from the soil surface. A general good agreement between observed and simulated volumetric water content data has been observed for both the calibration and validation period.

Their results together with others support the use of Hydrus 2D/3D commercial code to investigate this kind of agricultural-geotechnical problems and the use of sensors monitoring volumetric water content to calibrate and validate RWU models.

2_1_4 Experimental measurement of water potential distribution around root bulbs

Since the eighties, tensiometers have been used with multiple-purposes in the investigation of the phenomena related to the vadose zone. Tensiometers have become topic of research in the agricultural field because they can provide valuable information for an optimal water management of cultivated plants.

For example, Dirksen et al (1979) used 84 tensiometers around citrus trees to investigate regulated deficit irrigation and high frequency irrigation strategies that have the purpose of increase the water savings in the agricultural cultivation of these plants. The use of tensiometer information in order to control water availability in cultivated field has taken hold, f.e. Peterson et al (1992) designed an irrigation control valve mechanically linked to the soil water potential reading of a tensiometer. More recently, Bacci et al (2003) has investigated the adequacy of tensiometers to control irrigation in order to adapt irrigation schedule to the real plant exigencies. Koumanov et al (2006) investigated the spatial and temporal pattern of RWU of a microsprinkler-irrigated almond tree by means of catch cans, neutron probe and 8 tensiometers. Soil water dynamics were investigated under different irrigation conditions. Experimentally it has been observed that the zones of maximum root uptake were the same in both the irrigation conditions but no numerical simulations were carried out. Li and Liang (2019) designed a modified TDR probe able to simultaneously monitor soil water content and soil water potential in order to investigate the spatial and temporal variation of soil water dynamics underneath a root bulb of a tree, but such information have not been used to calibrate any numerical model simulating the observed phenomena. Many more examples could be made of tensiometers used in field to experimentally investigate soil water status near root bulbs (Sivapatham et al, 2000; Coelho et al, 2007), but no example of calibration and validation of hydraulic models that take into consideration RWU by means of water potential data are to the knowledge of the author.

2_2_1 Contribution of plant roots to shear strength

The interest in vegetation and its influence on slope stability started in the 1970s with laboratory tests, in situ tests and analytical models to compute the mechanical contribution of plant roots. One of the key mechanical characteristics of roots for geotechnical stability is their relatively high tensile strength, which contrasts with the weak tensile resistance of soils. Soil and roots, combined together, create a composite material similar to fibre-reinforced concrete. During soil shearing, shear stresses develop in the soil matrix and are transferred to the root fibres by interface friction along the root or by root tensile resistance (Baets

et al, 2008). The reinforcing effect depends on many parameters such as characteristics of the root system, root diameter distribution, density distribution with depth, root tensile strength, root tortuosity, orientation of the roots with respect to the principal directions of strain (Greenway, 1987).

Typical tests performed to assess the effect of roots on soil strength include laboratory direct shear device, in situ through pull out test or full-scale tests on vegetated slopes, some of which are described below. For what concerns analytical models to estimate the increase in soil shear strength due to roots, starting from the simple perpendicular root model of Wu (1979), more sophisticated models based on the deformational properties of the root reinforced soil have been elaborated.

2_2_1_1 The analytical model of soil-root interaction elaborated by Wu (1976)

Wu (1976), Waldron (1977) and Wu et al. (1979) were the pioneers who developed the first analytical models for root reinforcement of soils. The model proposed by Wu (1976) considers elastic roots in a soil matrix initially oriented normal to the shearing surface. Tangential forces are developed along the roots by friction or bonding between soil and roots. The model assumes that the soil friction angle is unaffected by the reinforcement and the tensile strength given by roots is fully mobilized during failure (all roots break at the same time). The enhancement of the strength is seen as an additional cohesion (apparent cohesion) introduced in the Mohr Coulomb (MC) failure criterion. This additional cohesion depends on the root area ratio (the ratio of the area occupied by roots over the area of the considered section, noted RAR) and the root tensile strength. The shear stress at failure s_f is the sum of the effective shear strength of the soil matrix given by MC criterion (s_s), the shear strength due to matric suction (s_Ψ) and the shear strength due to root system (s_r).

$$s_f = s_s + s_\Psi + s_r \qquad \text{[equation 35]}$$

It is generally accepted that root shear resistance is an additional cohesion (Abe and Ziemer, 1991; Belfiore and Urciuoli, 2004; Dias et al, 2017). According to Wu (1976), this term s_r could be expressed as follows:

$$s_r = (sin\theta + cos\theta\ tan\phi')t_r = k't_r \qquad \text{[equation 36]}$$

where t_r is the root tensile resistance per unit area of soil, ϕ' is the soil friction angle and θ the angle of shear distortion. Test results show that the quantity in round bracket, for an angle of distortion between 48°-72°, is about 1.2 (ranging between 1-1.3).

In Wu's model the tensile strength of root fibres per unit area of soil t_r is computed as the sum of the maximum tensile strength of each individual root. To simplify the calculation, roots are grouped in n classes and each class is identified by a different diameter range. In this way t_r is defined as the summation of the average tensile strength of each diameter class multiplied by RAR_n which is the root area ratio of class n. The root area ratio RAR is the area occupied by roots in a unit area of soil.

$$t_r = \sum_{n=1}^{N} t_{r,n}\, RAR_n \qquad \text{[equation 37]}$$

The global root reinforcement is computed inserting equation 37 in equation 36 and then in equation 35.

Looking at the stresses acting in the root within the shear zone (see Figure 49 from Dias et al, 2017), it can be seen that the root contribution to shear resistance is double: the normal force within the root (T_n) acts normally to the failure plane and increases the confining pressure in the shear band. In addition, the

tangential force within the root (T_t) directly resists shearing. On the contrary, during soil contraction, no tensions are created in the root fibre.

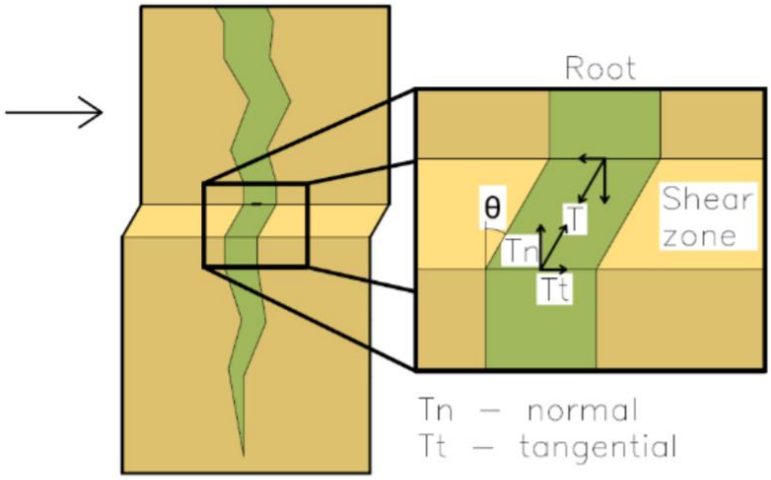


Figure 49: Stresses acting on a root during shear from Dias et al, 2017.

The main limit of Wu's model (1976) is the assumption that all roots break at the same time, which leads to an overestimation of soil reinforcement. Moreover in Wu's model pull out does not occur. What has been observed is that the failure of plant roots is progressive. When Wu's model is applied to an homogeneous root system the estimation results more accurate (Thomas and Pollen-Bankhead, 2010).

In a vegetated bank, roots are subjected to different forces: pull out in the crest due to tension cracks, pull out, breakage or buckling along the unstable slope and buckling due to compression at the slope toe (Schwarz et al, 2015; Dias et al, 2017). All the mechanisms have to be considered for a reliable estimation of the resistance provided by roots, while Wu's model considers just one failure mechanism.

Many models have followed Wu's model (1976) with main differences being relative to way roots fail. The main root breaking models are presented briefly below.

2_2_1_2 The Fibre Bundle Model (FBM)

Pollen and Simon (2005) introduced the Fibre Bundle Model (FBM) in which roots break progressively from the weakest to the strongest. Therefore, the maximum tensile strength is not the sum of the maximum strength of each root because they break at different instants depending on the way load is distributed. Different criteria of load distribution have been proposed by authors: based on root diameter, roots cross section area, equally distributed among the integer fibres or equally subdivided among all the fibres (more conservative in the estimation of the root reinforcement). Due to the fact that roots break in different moment, every time the load is redistributed among the integer ones according to two approaches: Local load sharing (LLS) and Global Load sharing (GLS). In GLS the load is redistributed among all the remaining roots, in LLS among the neighbouring roots. The latter is the most used because it requires a lower number of parameters. In the FBM, the root tensile resistance per unit area of soil t_r is computed as in equation 37 when the GLS approach is assumed. Roots are numbered from 1 to N, from the weakest to the strongest. RAR_j is the RAR of the weakest root j intact upon loading; j is the number of intact roots. So the root

tensile resistance t_r is equal to the max resistance of the weakest root multiplied by the root area ratio of root j multiplied by the number of intact roots.

$$t_r = \max(t_{r,j}\ RAR_j j) \quad \text{[equation 38]}$$

The global root reinforcement is computed inserting equation 38 in equation 36 and then in equation 35.

The limits of the FBL model are that root tortuosity is not considered, roots have all the same elastic properties and the majority of the roots break instead of being pulled out although, theoretically, both the failure mechanisms are possible (Dias et al, 2017).

2_2_1_3 Root Bundle Model (RBM)

Schwarz et al (2010) proposed the Root Bundle Model (RBM) which is an extension of the FBM approach, and also considers progressive root failure. The model presented by Schwarz is a stress-step loading and it is based on the relationship between pull-out force and root bundle elongation (displacement) and takes into consideration the root spatial distribution. More specifically RMS imposes equal displacement increments to a bundle of N roots. For each displacement increment the load taken by the N roots F_{tot} is the sum of the stresses in the same roots. When the stress exceed the maximum tensile strength of a root, the displacement increment is redistributed in the remaining intact roots. Forces are redistributed in each single root based on mechanical and geometrical properties The total force taken by the bundle of roots $F_{tot}(\Delta x)$ could de expressed as in equation 39.

$$F_{tot}(\Delta x) = \sum_{i=1}^{N} F_i(\Delta x) n_i \quad \text{[equation 39]}$$

Where Δx is the displacement increment, N is the number of diameter classes, n_i is the number of roots of diameter class i and $F_i(\Delta x)$ is the max pull out force of the class i.

2_2_1_4 Modified Root Bundle Model using Weilbull survival function (RBMw)

Schwarz et al (2013) proposed a modification of the RBM method to account for the variability of strength within roots of same diameter which belong to the same class using a statistical approach based on a Weilbull survival function. This function expresses the failure probability of a complex system beyond a threshold. The probability that a root survive as a function of displacement is formulated as follows:

$$S(\Delta x *) = \exp[-\left(\frac{\Delta x^*}{\lambda^*}\right)]^{\omega} \quad \text{[equation 40]}$$

Where ω is the Weilbull exponent (shape factor), λ^* is the scaling factor, Δx^* is the normalized displacement. Roots with the same diameter will fail at different displacement due to root-strength variability while roots with different increasing diameters will fail with increasing displacements.

The total force taken by the bundle of roots $F_{tot}(\Delta x)$ is obtained summing the force contributions from each root multiplied for the Weilbull survival function as follows:

$$F_{tot}(\Delta x) = \sum_{i=1}^{N} F_i(\Delta x)\ S(\Delta x *) \quad \text{[equation 41]}$$

Where N is the number of roots. To calibrate the model parameters data of laboratory tensile tests and in situ pull-out tests could be used.

2_2_2 Laboratory and in situ tests to characterise the mechanical response of roots and root-soil composite material

2_2_2_1 Root tensile strength t_r.

As highlighted before, root tensile strength is a crucial parameter to evaluate root reinforcement. Tests are often performed with standard load frames with measurement of force and displacement. The stress-strain curve is used to illustrate the tensile process of the plant root which is a natural elastic-plastic material. The mechanical response of roots depends on their water content (Baets et al, 2008; Yang et al, 2016) and, as such, roots should be tested at their natural water content. When tests are conducted weeks after excavating the plant, roots are conserved in alcohol solutions at a temperature around 4°C, this is necessary to preserve their tensile strength (Bischetti et al, 2003, De Baets et al, 2008). The explanation stays in the ability of the moist root of breaking beyond their elastic limit stress due to non-linear elastoplastic deformations while dry roots break at the stress limit. Like in many tensile tests on delicate material, clamping the specimen is the most delicate part of the test because the grips could damage the structure of the root causing their breakage close to the clamping positions, invalidating the result. Often to avoid these problems fine sand paper and rubber strips are used to clamp the roots without damage. Fine roots with diameter of few mm and roots with big diameters (>8mm) are the most problematic.

Many studies have demonstrated that root tensile strength is influenced by root characteristics (diameter, length tortuosity, orientation), site condition, season of the year and planting mode. It is relevant to highlight that among them the root diameter is the characteristic that affect tensile properties the most (Yang et al, 2016). Moreover, root tensile strength is strain rate sensitive , for example Cofie et al (2001) found that the tensile strength could increase up to 20% increasing the loading rate from 10 mm/min to 400 mm/min. It has been found also that the lignin and cellulose content in roots improve hardness and root elongation, respectively. Greater the root diameter, greater the percentage of cellulose and lower the percentage in lignin with consequential decline of the tensile strength (Zhang et al, 2014; Genet et al, 2005). Finer roots provide to the soil-root composite more ductility and ultimately an increase in shear strength due to their high resistance toward breakage (Genet et al, 2005). For all these characteristics, fine roots have a major role in withstanding surface erosion and runoff (Mickovski and Van Beek, 2009).

Many studies have stated that root diameters are negatively correlated to tensile strength (Baets et al, 2008; Genet el al, 2008; Tosi, 2007) and the general formulation is the following:

$$t_r = aD^{-b} \qquad \text{[equation 42]}$$

Where D is the mean root diameter and a and b constants.

In equation 43 the formulation of the power low of Bischetti et al (2003) is presented.

$$t_r = \frac{F_{max}}{\pi(\frac{D^2}{4})} \qquad \text{[equation 43]}$$

Where D^2 is the mean root diameter and F_{max} is the maximum force (N) needed to reach the root breakage. For this reason, during soil shearing, the finer roots with an high tensile strength slip while the thicker roots with a lower tensile strength break (Mickovski and Van Beek, 2009).

2_2_2_2 Pull-out test

Plants are frequently loaded mechanically from external forces such as wind, canopy weight or during slope deformation by moving soil. The consequence is that the base of the plant is pulled from the substratum. The mechanism with which the plant sustains a mechanical load is unknown because it depends largely on the morphology of the system and root properties.

The mechanics of plant anchorage is extremely complex and a well-established way to measure the anchorage resistance is the pull-out test. Such test is based on the measure of the peak force required to pull the plant root system from the soil and on the deformation at failure. During the test, performed at a constant rate, both displacement and load are recorded. Several researchers tried to correlate pull-out resistance to morphological characteristic of the root system investigated after the extraction from the soil. For example, Stokes et al (1996) noted that the branching angle of the lateral root system has a strong influence on the maximum pull-out force. Osman et al (2011) found that the max pull-out force increases as the stem diameter increases, as the shoot dry weight increases, as the Root Length Density (RLD) (total root length/soil volume) increases and it is much affected by root tensile strength. From all this, we can understand that the pull-out force is affected by the root system but also by the shoot profiles (e.g. shoot dry weight and stem diameter) even if, in a reduced way for the latter. Both shoot and root properties have to be always considered when dealing with pull-out test.

Pull-out tests are performed on in situ root bulb naturally grown but also in laboratory or in situ controlled experiments. In these cases, the pull-out force and displacement of single root or bundle of roots artificially infixed in the soil are investigated. For example Schwarz et al (2011) tested straight roots, tortuous roots, roots with and without branching points, roots at different water content in order to investigate how all these aspects affect the maximum pull-out force and displacement at the maximum pull-out force.

2_2_2_3 Laboratory shear tests

To quantify the contribution of roots to soil shear strength in a controlled environment, shear tests are often conducted in laboratory (direct shear tests or triaxial tests). Many publications have highlighted the significant increase in soil shear strength imputable to roots (Day, 1994; Campbell and Hawkins, 2003, Ali and Osman, 2007, among others). Experimental results show that roots enhance soil shear strength through an increase in cohesion together with the internal friction angle. Nevertheless, It is important to highlight that the major contribution of the roots to shear strength is due to the increase in the cohesive intercept as found by many authors (Waldron, 1977; Wu and Watson, 1998, Veylon et al, 2015). The increase in the friction angle is of very small entity for coarse-grained soil (sand) (and often null) while it assumes a relatively higher relevance in fine-grained soils (Jiang et al, 2010). This is due to the fact that in order to register a change in the soil internal angle important particle arrangements on the shear surface are necessary. Studies have shown that coarse materials undergo minor rearrangements due to roots on the shear surface that act like a barrier to soil particles. Due to the more cohesive nature of fine-grained material, particles could rearrange easily around roots upon shearing especially when the soil has an high water content causing a lubrification of the interfaces between root and soil particles.

The quantity of roots crossing the failure surface has an high relevance too: the smaller the quantity of roots, the smaller the particle rearrangement and the influence on the internal friction angle. The same conclusions have been obtained using natural fibres and synthetic fibres (polypropylene, polyethylene and nylon) as done by Gray and Ohashi (1983); Hejazi et al (2012).

It has been observed also that roots tend to contribute more to an increase of the soil shear strength when high moisture content is present (Coppin and Richards, 2007).

The presence of roots on the shear surface causes a delaying of the soil deformation upon shearing (deformation occurs at a slower rate) which affects in turn the resistance peaks and produces an increment of the cohesion as observed experimentally (Maffra et al, 2019). So, when roots cross the failure surface an increase in the cohesion parameter is expected. Maffra et al (2019) observed that upon shearing (direct shear tests) the vertical deformations are positively influenced by roots presence in fact soil+roots composite undergoes smaller deformations by compression indicating that roots can help in withstanding vertical stresses both for fine and coarse grained materials.

Interestingly, many root-soil specimens tested in the laboratory have been reconstructed with roots packed in the soil along a preferential direction (horizontally, vertically or both) in order to investigate the influence of root density or orientation (Zhang et al, 2010; Liu et al, 2011, Mickovski et al, 2009). Similarly, fibres analogous to natural roots have been employed to avoid the natural variability in the root system and in order to have a greater control on the sample (Wu et al, 1988). If in one hand, when an ad hoc sample is tested, a great control over sample properties is possible, on the other the sample does not represent the natural processes that occur in nature. In particular the friction at the root-soil interface, the tortuosity of the root, the adhesion between root and soil, the differences in the properties of the rhizosphere (narrow region around roots interested by root secretion and microorganism life), root anchorage. A very restrict number of studies has grown plants in the sample prior of their testing in a direct shear device (Waldron and Dakessian, 1982; Waldron et al., 1982; Frydman and Operstein, 2000) or prior of their testing in a triaxial apparatus (Miranda and Mahler, 2017).

2_2_2_4 In situ shear tests

In such tests, a soil block is subjected to a direct shear by means of a system similar to the laboratory direct shear box. Initially designed to test in situ rocks, in situ shear tests have been adapted to soils and composite root-soil materials. From the pioneer work of Endo and Tsuruta (1969), the shear in situ device has been modified multiple times by Ziemer (1981), Wu et al (1988); O'Loughlin (1981) among others. The common practice to operate a in situ-shear test is to excavate the soil around a test block that could be for example a plant whose external part has been cut in order to maintain only the below ground structure. A typical device is composed of a four-sided metal box containing the soil block with open upper and bottom parts. The base of the block simulates the potential failure surface. A horizontal force is applied to the base by means of hydraulic jacks acting against the two walls of the excavation trench and it is measured by means of a pressure transducer that records the hydraulic pressure. Displacements are usually measured tracking the positions of pins driven in the soil surface and the position of the visible roots using a series of photographs. Shear deformations of in-situ tests are not constrained to a thin zone (as in laboratory direct shear tests). Because the thickness of the shear zone depends on the amount of reinforcement, in situ shear tests are able to simulate the condition of a thick shear zone as the one that could be formed from a composite layer saturated from the top down (Wu and Watson, 1988).

2_2 INVERSE MODELLING AND PERFORMANCE ESTIMATION OF THE INDIRECT SIMULATION

Modelling soil water dynamics requires the definition and calibration of numerous parameters associated to the definition of the SWRC and of the hydraulic conductivity function. These parameters are obtained usually in laboratory but a problem of representativeness at the field scale due to limitation of core sample size is often encountered (Le Bourgeois et al, 2016). Indirect methods such as inverse modelling are succeeding both in large scale studies and in the small scale laboratory experiment because they give efficient estimations and they are easy to set up and process.

Inverse analysis is based on the definition of an objective function which measures the agreement between simulated and observed datasets. The Levenberg-Marquardt (L-M) algorithm tries to minimize the objective function to perform inverse estimation of hydraulic parameters in steady state or transient flow models. The L-M algorithm needs an initial estimate of the unknown parameters that are going to be optimized and a dataset of observation points in order to measure the discrepancy between the observed values and the predicted system response (Simunek et al, 1998; Hopmans et al, 2002; Simunek et al, 2012a; Simunek and Hopmans, 2012, Simunek et al, 2012b). The general formulation of the objective function is the following:

$$\phi(\beta) = \sum_{j=1}^{m_q} v_j \sum_{i=1}^{n_q} w_{i,j}[(q_j^*(x,t) - q_j(x,t,\beta)]^2 \qquad \text{[equation 44]}$$

where β is the vector of the optimized parameters, $q_j^*(x,t)$ are the measured variables, $q_j(x,t,\beta)$ are the simulated variables, $w_{i,j}$ the weights attributed to the measured variables, m_q are the different measurement types (water content, water pressure head, cumulative flux etc), n_q are the number of data for each measurement type j, v_j is a scaling factor which takes into consideration the different typologies of measurements considered. More information could be found in chapter §2_7 and subchapter §2_7_1 (PART I) of the present thesis.

When a large pool of indirect simulations is performed, a set of metrics/indices could be used to evaluate in a quantitative way the simulation output in order to individuate the set of optimized parameters which enhances the most the model performance. In Table 62, a summary of the analytical expressions of the metrics/indices used in the following chapters is presented.

ACCURACY METRICS			
Metric	**Mathematical formula**	**Description**	**Best value**
MAE	$MAE = \frac{1}{n} \cdot \sum_{1}^{n} \lvert M_i - S_i \rvert$	*Mean Absolute Error*	0
MSE	$MSE = \frac{1}{n} \cdot \sum_{1}^{n} (M_i - S_i)^2$	*Mean Square Error*	0
RMSE	$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{1}^{n} (M_i - S_i)^2}$	*Root Mean Square Error*	0

CRM	$CRM = \frac{(\sum_1^n M_i - \sum_1^n S_i)}{\sum_1^n M_i}$	*Coefficient of Residual Mass*	0
R^2	$r^2 = \frac{(\sum w_i M_i S_i - (\sum w_i M_i * \sum w_i S_i)/\sum w_i)^2}{(w_i M_i^2 - \frac{(\sum w_i M_i)^2}{\sum w_i})\,(w_i S_i^2 - \frac{(\sum w_i S_i)^2}{\sum w_i})}$	*Coefficient of determination considering the internal weights*	1
EFFICIENCY METRICS			
Metric	**Mathematical formula**	**Description**	**Best value**
NSE	$NSE = 1 - \frac{\sum_1^n (M_i - S_i)^2}{\sum_1^n (M_i - \bar{M})^2}$	*Nash-Sutcliffe efficiency*	1
IA	$IA = 1 - \frac{\sum_1^n (M_i - S_i)^2}{\sum_1^n (\lvert S_i - \bar{M}\rvert + \lvert M_i - \bar{M}\rvert)^2}$	*Index of Agreement*	1
NSE_j	$NSE_j = 1 - \frac{\sum_1^n \lvert M_i - S_i\rvert^j}{\sum_1^n \lvert M_i - \bar{M}\rvert^j} \quad j \in N$	*Modified efficiency*	0
IA_j	$IA_j = 1 - \frac{\sum_1^n \lvert M_i - S_i\rvert^j}{\sum_1^n (\lvert S_i - \bar{M}\rvert + \lvert M_i - \bar{M}\rvert\ )^j} \quad j \in N$	*Modified Index of Agreement*	1
NSE_{rel}	$NSE_{rel} = 1 - \frac{\sum_1^n \left(\frac{M_i - S_i}{M_i}\right)^2}{\sum_1^n \left(\frac{M_i - \bar{M}}{\bar{M}}\right)^2}$	*Relative efficiency*	1
IA_{rel}	$IA_{rel} = 1 - \frac{\sum_1^n \left(\frac{M_i - S_i}{M_i}\right)^2}{\sum_1^n \left(\frac{\lvert S_i - \bar{M}\rvert + \lvert M_i - \bar{M}\rvert}{\bar{M}}\ \right)^2}$	*Relative index of agreement*	1
$IA_{j=1}$	$IA_{j=1} = 1. - \frac{\sum \lvert M_i - S_i\rvert}{\sum (\lvert S_i - \bar{M}\rvert + \lvert M_i - \bar{M}\rvert)}$	*Modified index of agreement with j=1*	1
IA'	$IA' = 1. - \frac{\sum \lvert M_i - S_i\rvert}{2 * \sum \lvert M_i - \bar{M}\rvert}$	*Modified index of agreement*	1

Table 62: Metrics/Indices used to quantify the departure of the output of the indirect simulation from observed and experimental measurements. This pool of metrics/indices are used in the first phase of the calibration process of the model parameters.

Where M_i is the observation value, S_i is the simulated value, n the number of available real data, w_i the weight associated to each measurement of a dataset; $\bar{M}$ the mean of the observed values and $\bar{S}$ the mean of the simulated values.

The following parameters/coefficients in Table 63 are used in the second phase of the calibration process to analyse the results of the inverse analysis (last iteration of the indirect simulation) and to evaluate the accuracy of the prediction of the optimized parameters:

Sum of Squared Residuals	**SSQ**	The sum of the squares of the residuals (deviation between observed-simulated values) is a good indicator because it measures the discrepancy between data and model estimation from a global point of view (all observation points considered together with their internal weights).
Mass Balance Error	**MBE**	Mass Balance Error is an indication of the accuracy with which the finite difference matrix equations have been solved. A mass balance error equal to 0% means that the mass is perfectly conserved and the matrix solver is solving the equation with very high accuracy in the previsions. It is always desirable to have a Mass Balance Error below 1%.
Standard Error Coefficient	**SEC**	The Standard Error is an estimate of the standard deviation of the estimated coefficient (parameter). It could be useful to estimate how much the coefficient varies across different cases. Lower the Standard Error Coefficient, greater the precision with which the regression coefficient is measured.
Number of iterations performed in an indirect simulation (NoI)	**NoI**	In each iteration of the inverse analysis, the matrix solver solves the governing equations using updated parameters, so with each and new iteration the set of optimized parameters becomes increasingly different from the initial one.
Confidence Interval of an optimized parameter	**CI**	A Confidence Interval suggests the uncertainties of a particular statistic, in this case the estimation of the optimized parameter. It indicates the range of possible values that could include the true value of a considered population with a certain degree of confidence (usually set to 95%).
Pearson Correlation Coefficient of a couple of optimized parameters	**CC**	It is a statistic measure of the linear correlation between two variables, in this case, a couple of optimized parameters. It has a value between +1 and -1 (+1 is a total positive linear correlation, 0 no linear correlation and -1 is a total negative linear correlation). It is defined as the covariance of the couple of variables divided by the standard deviation of each variable multiplied by each other.

Table 63: Parameters/coefficients used to analyse the results of the inverse analysis (second phase of the calibration process)

3 THE LARGE-SCALE EVAPOTRANSPIRATION EXPERIMENT

3_1 MATERIAL CHARACTERIZATION

3_1_1 Original in situ soil

About 1m^3 of soil was collected from the site and its basic geotechnical properties were determined in laboratory (see).

Table 64 Main physical properties of the in situ soil obtained from laboratory testing.

	Average value	Standard deviation	n° of measurements		Grain size distribution
Organic content (%)	4.611	0.203	5	**% gravel**	0
G_s	2.575	0.011	5	**% sand**	26.5
Liquid limit (%)	51.81	1.55	3	**% silt**	47.3
Plastic limit (%)	27.5	0.3	5	**% clay**	26.2
PI	24.3			**C_u**	26.5
Ks (m/sec)	5.57E-08	1.56E-08	3	**C_c**	0.6

According to the Australian Standard AS1726:2017, the original soil is an organic sandy clay, at the boundary between medium plasticity and high plasticity (organic clay of high plasticity). The fine fraction constitutes 73% of the soil with a coarse sand making another 26.5%. The rest is organic matter. The grain size distribution, obtained by a combination of sieving (AS 1289.3.6.1-2009) and sedimentation with an automated sediGraph (Micrometrics SediGraph III), is shown in Figure 50. The sedigraph uses an X-ray beam to detect changes in the concentration of a suspended sediment. The specific weight of the soil particles has been obtained using the gas pycnometry technique (AccuPyc II micrometrics) with helium.

pH measurements showed that the soil is extremely acidic with values ranging from 4.8 to 5.4 creating a hostile and undesirable environment for vegetation. A high level of hydrogen, which causes the acidity, results in high concentrations of aluminum, a substance that is toxic for plants. The electrical conductivity of the soil, however, revealed a low salinity, which is desirable for plant growth.

The high soil acidity and its low saturated permeability (determined by a falling head permeameter) led to the decision to improve the soil as follows:

- Lime was added to the mixed soil to the rate of the 0.04% of the dry mass of the soil in order to increase the pH to ~ 6.
- The soil has been mixed with 65% in mass of a clean coarse river sand in order to increase the saturated permeability and reduce the air entry value of the soil. This modification is driven by the fact that roots need access to oxygen for the plant to live, which is difficult to achieve with a high plasticity clay.

The grain size distribution of the sand is presented in Figure 51. The sand contains about 25.44% of gravel, 74.56% of sand and 0% of fine fraction. Its D_{10} , D_{30} and D_{60} diameters are ~0.3 mm, ~0.63 mm, ~1.2 mm, respectively. With a uniformity coefficient C_u of ~4.0 and a coefficient of curvature C_c of ~ 1.8, the sand is well-graded.

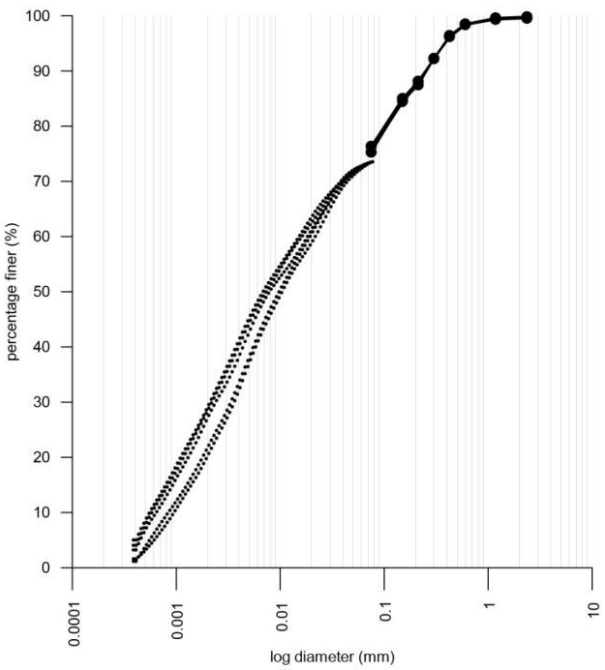


Figure 50 Grain size distributions of the investigated soil. The coarse material has been tested by sieving (according to AS 1289.3.6.1-2009) while the finer material by SediGraph method (Micrometrics SediGraph III).

The improved soil will be used for large scale evapo-transpiration experiment but also in situ when planting the seedlings of Melaleuca Styphelioides. It is recommended to excavate a small volume of original soil and replace it by the improved soil.

The rest of this section will provide the properties of the mixed soil and the potting mix in which the Melauleuca Styphelioides seedlings grew.

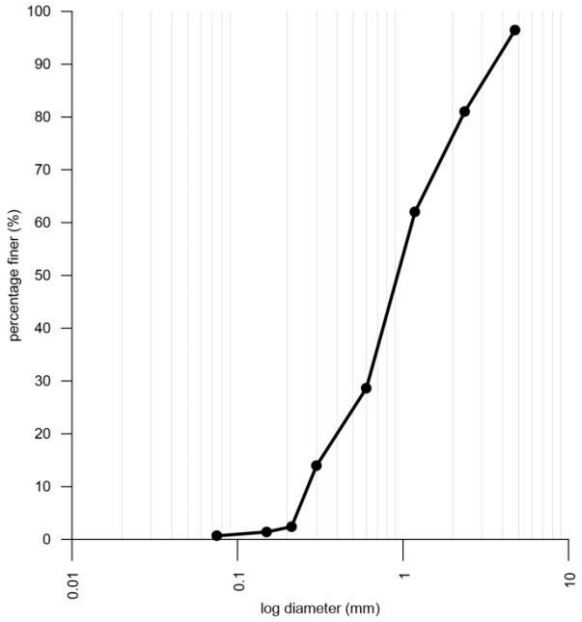


Figure 51: Grain size distributions of the coarse sand added to the original sandy clay for the large -scale evapotranspiration experiment.

3_1_2 Improved soil

3_1_2_1 Basic Geotechnical Properties

As for the original soils, basic geotechnical properties were measured (see Table 63). The specific weight of the soil is equal to 2.623 (standard deviation 0.008). The organic content of the mixed soil is equal to 3.94% (standard deviation of 1.9%). The liquid limit has an average value of 26.6% (standard deviation 1.52%) and the plastic limit has an average value of 22.5% (standard deviation 0.004%), the plasticity index PI is equal to 4.1.

Table 65: Main physical properties of the mixed soil (top soil + coarse sand) used for the laboratory tests.

	Average value	Standard deviation		Grain size distribution
Organic content	3.94	1.90	**% gravel**	20.0
G_s	2.623	0.008	**% sand**	45.0
Liquid limit (%)	26.6	1.52	**% silt**	22.5
Plastic limit (%)	22.5	0.004	**% clay**	12.5
PI	4.1			

In addition, some proctor compaction tests were conducted to identify the optimum moisture content (14.7%) and maximum dry density (2.10 $Mg.m^3$).

3_1_2_2 Soil water retention curves and microstructure

The soil water retention curve was determined both in terms of gravimetric water content and saturation degree (these will be referred to as “gravimetric retention curve” and “volumetric retention curve”). The procedure followed to get the gravimetric retention curve is described below:

- Water was added to the dry soil (proportion of 1.5 times the liquid limit) and an ultrasonic probe was used to mix it homogeneously ready for the next phase.
- The mix soil+water was then placed in an oedometer cell and consolidated under small vertical stress to avoid the material being squeezed outside the cell. The vertical stress was progressively increased to reach a final void ratio of 0.8. This void ratio was targeted because it coincides with the void ratio of the soil placed in the large scale apparatus.
- The cylindrical soil sample was removed from the oedometer ring and placed in a WP4 Dewpoint potentiometer airtight plastic container.
- Suction was measured directly using a high capacity tensiometer (up to 1MPa, accuracy of +/- 10kPa) on the sample surface and by placing the WP4 cup in the dewpoint potentiometer for value of suctions above 1MPa. This technique allows a very low disturbance of the sample throughout the whole test.
- To obtain the whole retention curve, the sample was incrementally dried after each suction measurement and left to equilibrate for at least two days in a sealed vacuum bag before record another suction measurement. This means that the SWRC is not at constant void ratio and for this reason the total change in the degree of saturation includes changes due to suction and due to void ratio. Despite that, many authors (Lloret and Villar 2007; Romero et al, 1999; Imbert et al, 2005; Sun et al, 2010 among others) have found that void ratio has a minor influence on the relation between water content and suction, especially moving to high suction value.

Figure 52 shows the gravimetric retention curve where two datasets obtained from the same soil (e=0.8) are superimposed, to assess repeatability of measurement.

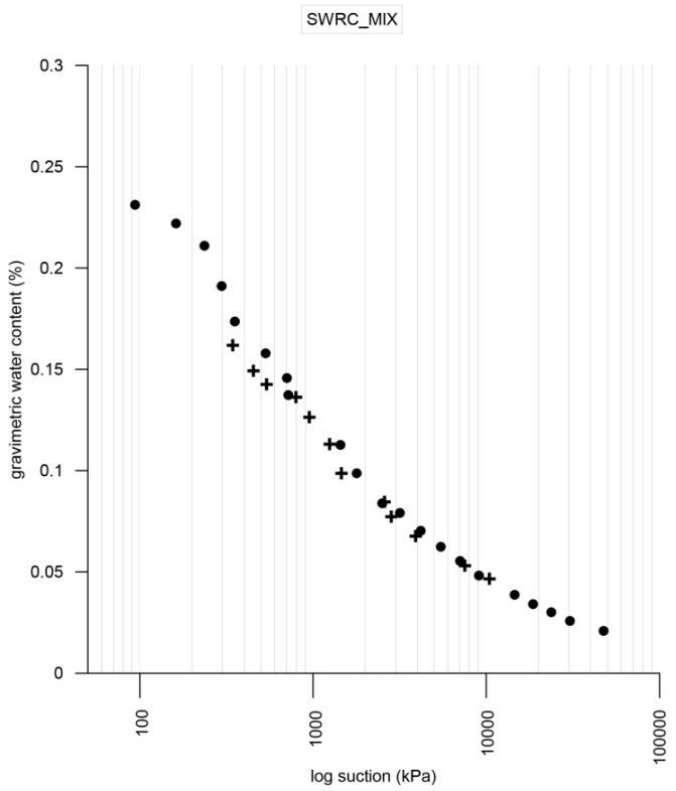


Figure 52 Retention curve of the improved soil in terms of gravimetric water content.

To express the retention curve in terms of volumetric water content and degree of saturation S_r, it is necessary to track how the soil shrinks as it dries (referred to as a shrinkage curve). This was achieved using the hand spray plaster method developed by Liu and Buzzi (2014) following the steps described below:

- A consolidated soil sample (as per preparation for the gravimetric SWRC) was extruded from the metal ring and cut into a $1cm^3$ cube. Offcuts were used to measure gravimetric water content. The soil cube was weighed (photo E in Figure 53).
- The soil cube was sprayed with a layer of flexible plaster (Elastoplast Spray Plaster) which is waterproof and permeable to water vapour.
- The cube and the plaster layer were let to set in a fog room (to minimize the evaporation), after which the cube was immersed into silicon oil and its volume measured using the displacement technique (photo G, Figure 53). The plaster coating prevents the intrusion of the oil into the soil cube.
- After the weighting, the cube was cleaned from the oil on the surface (image H, Figure 53) and dried for a brief period and left to equilibrate for two days before another measurement.
- The process was repeated until the full shrinkage curve was produced.

The soil shrinkage curve can then be combined to the gravimetric retention curve in order to produce the retention curves in terms of volumetric moisture content and saturation degree (see Figure 54).

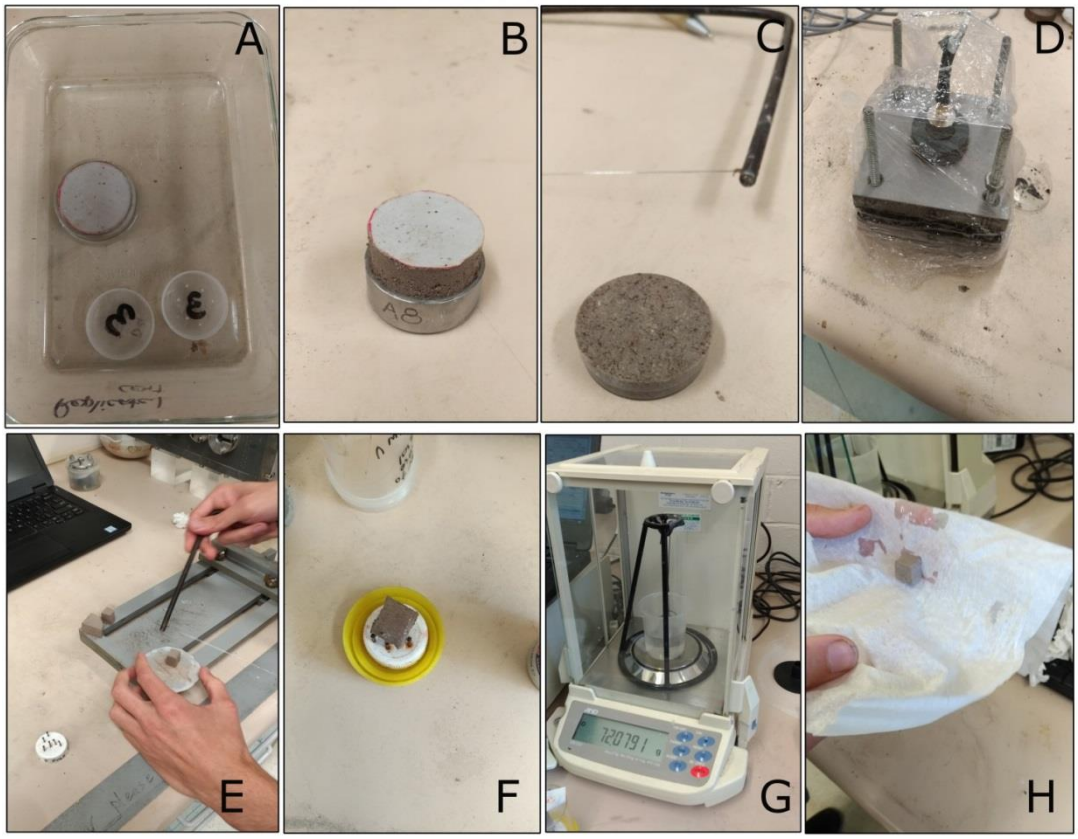


Figure 53: In photo A the sample created in the oedometer cell with the required void ratio in saturated conditions; in photo B the extrusion of the sample from the oedometer ring, in photo C the sample inserted in the WP4 container with holes on the bottom in order to avoid the entrapment of air, in photo D the tensiometer during the suction reading on the surface of the sample (plastic film avoid the loss of water during the reading). In photo E the creation of a small cube from the sample created in the oedometer cell, in photo F the cube sprayed with a layer of flexible plaster, in photo G the soil cube during the hydrostatic weight in silicon oil, in photo H the soil cube during the cleaning of its sides from the oil after the weighting.

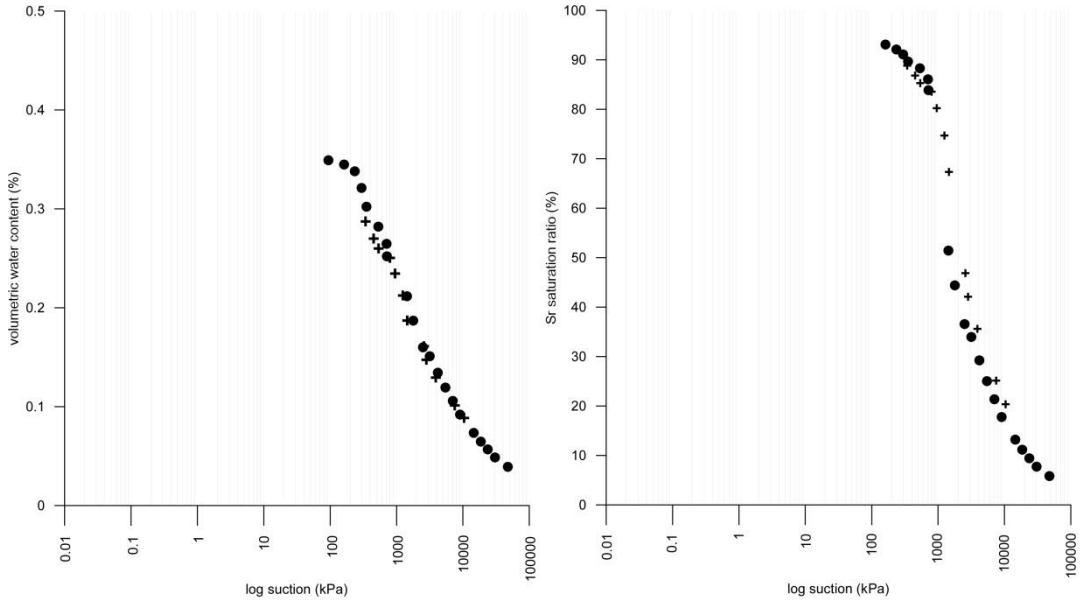


Figure 54: On the left side the SWRC of the mixed soil in terms of volumetric water content (vwc) vs suction (kPa) and on the right side the SWRC in terms of saturation ratio (Sr%) vs suction (kPa)

The microstructure of the soil has been characterised by mercury intrusion porosimetry (MIP) using a Micrometrics AutoPore IV. Before conducting the MIP analysis, the soil (initial moisture content around 10% under ambient humidity) was quasi-statically compacted to a void ratio of 0.8 and small pieces of compacted soil (size of about 5 cm^3) were placed in a freeze dryer to remove the water while preserving the soil structure (as per Yuan et al., 2016).

Figure 55 shows a bimodal distribution, which is consistent with a soil compacted dry of optimum (Russo et al, 2016). The dominant pore size is at about 0.07 mm. The small peak associated with micropores suggests that the total volume of intra-aggregate pores is considerably smaller that the volume of inter-aggregate pores and it is expected that the behaviour of the compacted improved soil will be governed by macropores.

The pore size distribution can be used to estimate the air entry value of a soil Ψ (kPa) by applying Young-Laplace equation to the dominant pore size and using the water contact angle (Washburn, 1921):

$$\Psi = \frac{26cos\alpha}{R} \qquad \text{[equation 45]}$$

Where R is the radius of the dominant pore size (here 0.07mm), α=0 (contact angle between water and surface of the solid material tested) and $\mathcal{G}$=0.075 kPa mm (surface tension) (Watabe et al, 2000).

This equation assumes a cylindrical-shaped pore. According to Young-Laplace equation, the air entry value of the improved soil compacted at e=0.8 is about 2kPa. The relatively large volume of macropores is expected to facilitate root growth.

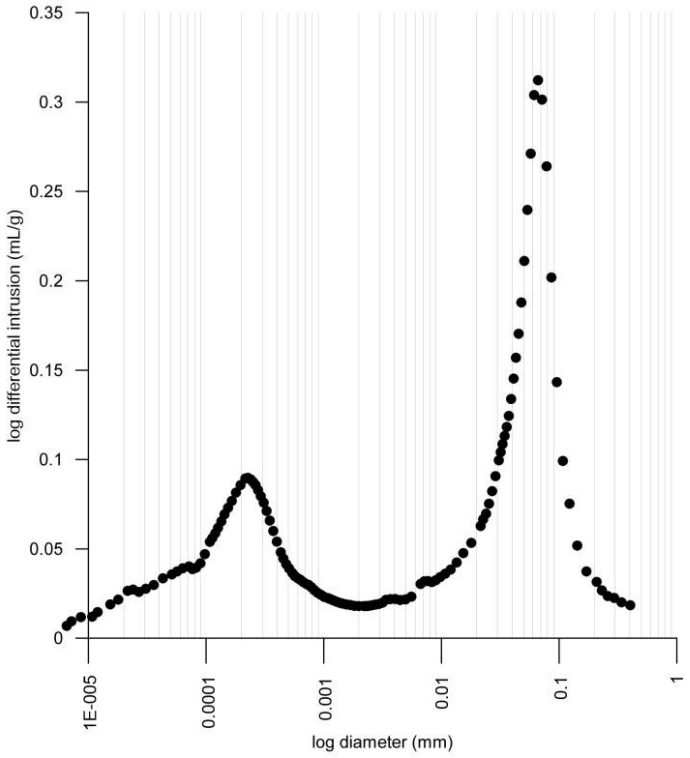


Figure 55: Pore size distribution by mercury intrusion porosimetry (MIP) for the investigated mixed soil.

The RETC free computer program developed by van Genutchen et al (1991) is a non-linear least-squares optimization program used to fit analytical function to observed water retention and hydraulic conductivity data starting from an initial set of parameters. The program iteratively changes the parameters of the analytical function trying to minimize the residuals between observed-simulated datasets.

RETC has been used to analyse the gravimetric water content data vs suction data (Figure 52) and estimate a soil water retention curve using the van Genutchen model (acronym VGM). RETC program allows also to estimate the unsaturated hydraulic conductivity function using the theoretical pore-size distribution model of Mualem (Mualem, 1976) assuming to know one observed conductivity value (not necessary at saturation).

The air entry value of the macropores obtained by mercury porosimetry (~2kPa) has been used as one of the initial parameters (in the zero iteration), and its value has been fixed over the iterative process while the rest of the hydraulic parameters (θ_s, θ_r, n, α, l) are let free to change in order to reduce the residuals (observed-simulated SWRC).

The VGM and the Mualem's hydraulic conductivity functions have been extensively presented in chapter §2_3 (PART I) and here the equations are briefly recalled.

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1+|\alpha h|^n]^m} \quad h<0 \qquad \text{[equation 46]}$$

$$\theta(h) = \theta_s \quad h\geq 0 \qquad \text{[equation 47]}$$

$$m = 1 - \frac{1}{n} \quad n > 1 \qquad \text{[equation 48]}$$

$$K_s(h) = K_s S_e^l [1 - (1 - S_e^{1/m})^m]^2 \qquad \text{[equation 49]}$$

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} \qquad \text{[equation 50]}$$

where θ_s is the saturated water content, θ_r the residual water content, α the inverse of the air entry value, n is a pore-size distribution index, S_e the effective water content.

Only the drying branch of the SWRC has been investigated (and its hydraulic parameters individuated) because the numerical FE model that has been elaborated from the large-scale apparatus experiment adopts the VGM and this choice implies to not consider the soil hysteretic effects (and the hydraulic parameters of the wetting branch). The VGM needs a reduced number of parameters allowing to focus the analysis on the plant evapo-transpiration contribution and the parameters of the root water uptake distribution.

Table 66 presents the fitted hydraulic parameters of the VGM describing the drying branch of the SWRC of the consolidated soil. Figure 56 presents the experimental points and the fitting VGM curve.

Table 66: Calibrated parameters of the van Genutchen model for the retention curve of the improved soil obtained from laboratory data (gravimetric water content vs suction).

	θr (-)	θs (-)	α (1/m)	n (-)	l (-)
Improved soil (VGM)	0.0001	0.45	4.26	1.422	0.5

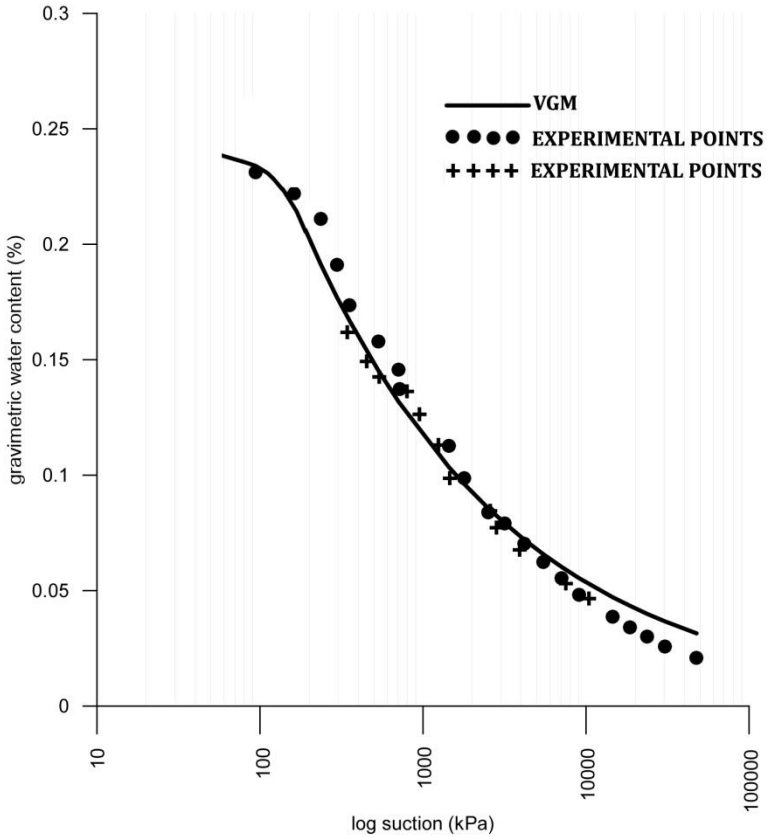


Figure 56: The experimental points and the fitting VGM curve are superimposed in the same graph.

3_1_2_3 Saturated Permeability

The saturated permeability was measured on specimens of improved soil, compacted at a void ratio of 0.8 and at a water content of 5% (air dried). The test was conducted in a Rowe Cell by imposing a positive water pressure (2kPa) at the lower part of the compacted soil specimen using a GDS pressure controller. Water was collected at the top end of the cell and weighted using a precision scale (see Figure 57). The permeability was computed from the value of the outflow rate in steady state conditions. A value of $2.93 \cdot 10^{-7}$ m/s was found.

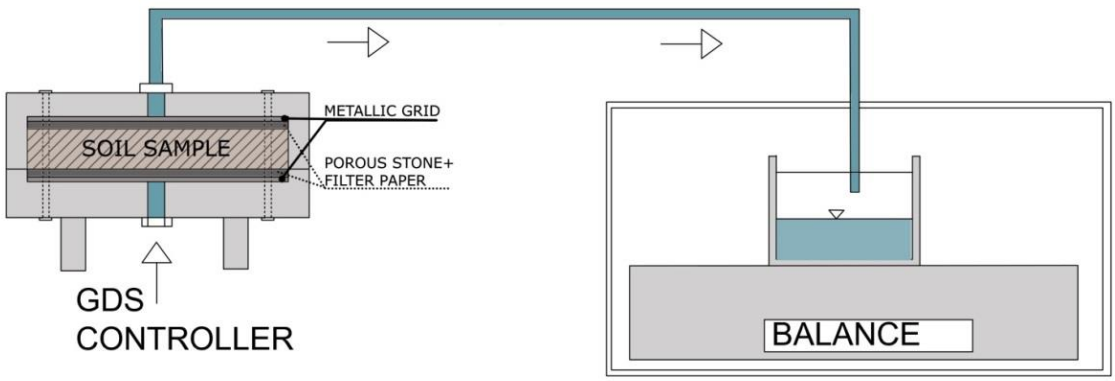


Figure 57: Sketch of the experimental set up (Rowe Cell) used to measure the saturated permeability of the improved soil (compacted at a void ratio of 0.8).

3_1_2_4 Unsaturated permeability: the experimental infiltration column

The permeability function of the unsaturated improved soil was estimated using the transient profile method which requires (1) an experimental set up where the progress of a wetting front is monitored and (2) a numerical model which replicates the experimental observations and in which the permeability function is progressively tuned until the numerical results match the experimental observations (Watson et al, 1966; Naasz et al, 2005; Leung et al, 2016).

Figure 58 shows the experimental setup (the column) used for a 1D steady state flow. The column has a rectangular section and is 384 mm high. The soil is placed in it in 35 mm layers with an initial water content of 6.5% in order to avoid tensiometers cavitation, then layers are progressively compacted using the same load each time to achieve a consistent void ratio of 0.8 across the whole specimen (see photo B in Figure 58). Water is fed at the bottom of the column at a constant rate of (6.7 mm^3/sec) using a GDS controller. The bottom of the column contains a porous stone to distribute the water and generate a 1D flow without much edge. The top of the column is open to the atmospheric pressure but covered with plastic film in order to avoid water loss from the soil sample during the test. 6 holes on the side of the column receive the high-capacity tensiometers used to monitor suction changes during the experiment. Here, for the presented transient profile experiment, only two tensiometers were used, and these were placed at the first and second holes (95 mm and 155 mm from the base of the column).

3_1_2_5 Infiltration column model

The 1D flow in the column was modelled using the commercial code Hydrus 2D by Pc-Progress. The objective is to numerically replicate the experimental observations of temporal changes in suction and in moisture content (inferred from the SWRC and suction measurements) within the column. The numerically calibrated permeability function will then be taken as the permeability function of the soil. Due to the fact that the soil column is subjected to an advancing wetting front and initially soil has water content close to zero, the experiment carried out investigates the wetting branch of the SWRC.

The simplified geometry and the position of the observation points can be visualized in images D and E. A “no flux” boundary condition was applied to the top and all side surfaces of the model while a constant flux was set for the bottom surface (see image F in Figure 59). The minimum time step was 0.01 sec together with a pressure head tolerance of 0.001 m and a water content tolerance of 0.001 (-).

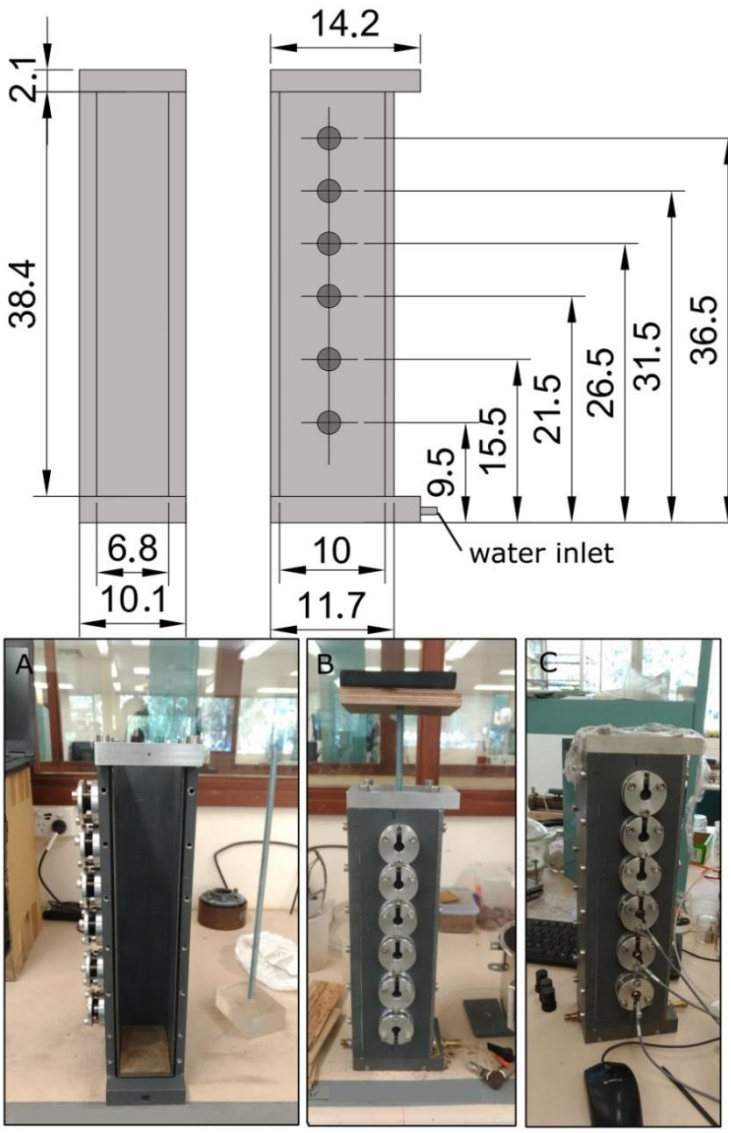


Figure 58: Top sketch shows the geometry of the column (all dimension in cm); photos A,B,C of the column during the setup of the test.

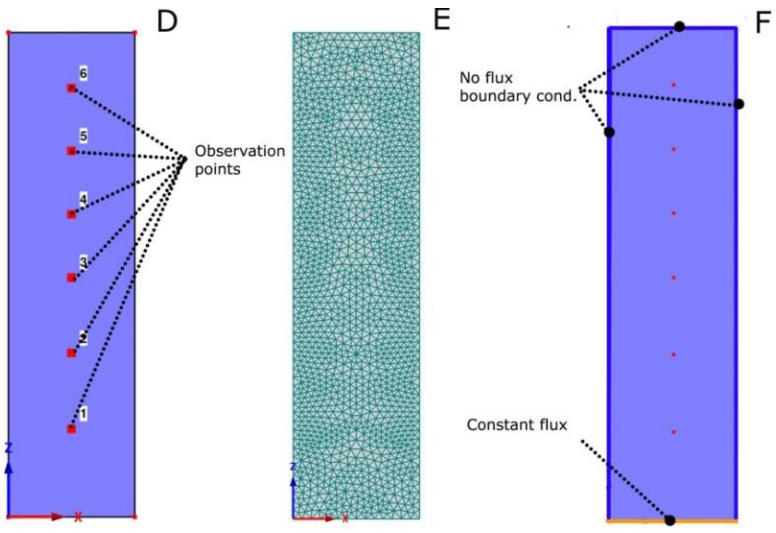


Figure 59: images D, E, F show how the experiment in the infiltration column has been modelled using Hydrus 2D.

3_1_2_6 Sensitivity analysis to perform calibration

A sensitivity analysis has been performed before calibrating the hydraulic parameters in order to identify the parameters that have the highest influence on the numerical results. The sensitivity analysis is based on coefficients which represent the change in the model output due to an infinitesimal change of the parameter: one at a time, each hydraulic parameter of the investigated soil is varied of +/-1%. When an input parameter was changed from its initial value to its varied value, the effect of that change on the numerical results was assessed in terms of sensitivity coefficient. This latter parameter was calculated as the initial numerical output minus the varied numerical output divided by the initial numerical output as in equation 183. The numerical output used is the pressure head at observation point 2.

$$S(t, b_j) = \frac{|Y(b_j + \Delta b_j e_j) - Y(b_j)|}{Y(b_j)} \qquad \text{[equation 51]}$$

where $S(t, b_j)$ is the change in the variable Y corresponding to a 1% change in the parameter b_j; Δb_j=0,01 b_j; e_j is the j_{th} unit vector. The parameters that are investigated in the present sensitivity analysis are the VGM and Mualem's parameters (θ_r, θ_s, α, n, K_s, l) presented in equations from 46 to 50.

Table 69 summarizes the outcomes of the sensitivity analysis. It can be observed that:

- the model output is insensitive to variations of residual water content θ_r (sensitivity coefficient in negative sign of -0.05 %). Consequently, the residual water content will be excluded from the calibration procedure.
- the highest model sensitivity pertains to the saturated water content θ_s (sensitivity coefficient in negative sign of -30%)
- the model is also very sensitive to the shape parameter n (sensitivity in negative sign coefficient of -13%).
- the model shows a low sensitivity for the K_s parameter (sensitivity in negative sign -6.6%)
- the model is almost insensible to the l parameter (sensitivity coefficient in negative sign -2.7%)

Variation of input parameter [%]	l	θ_r	θ_s	α	n	K_s
+1%	-2.86	-0.05	-29.95	-8.68	-12.97	-6.09
-1%	-2.69	-0.04	-23.19	-7.93	-13.26	6.62

Table 67:Summary of the sensitivity analysis on the 1D infiltration column model. The table reports the min values (in negative sign) of the sensitivity coefficient measured as a result of a change of the input parameter by +/- 1%. The input parameters considered are l, θr, θs,α,K_s,n.

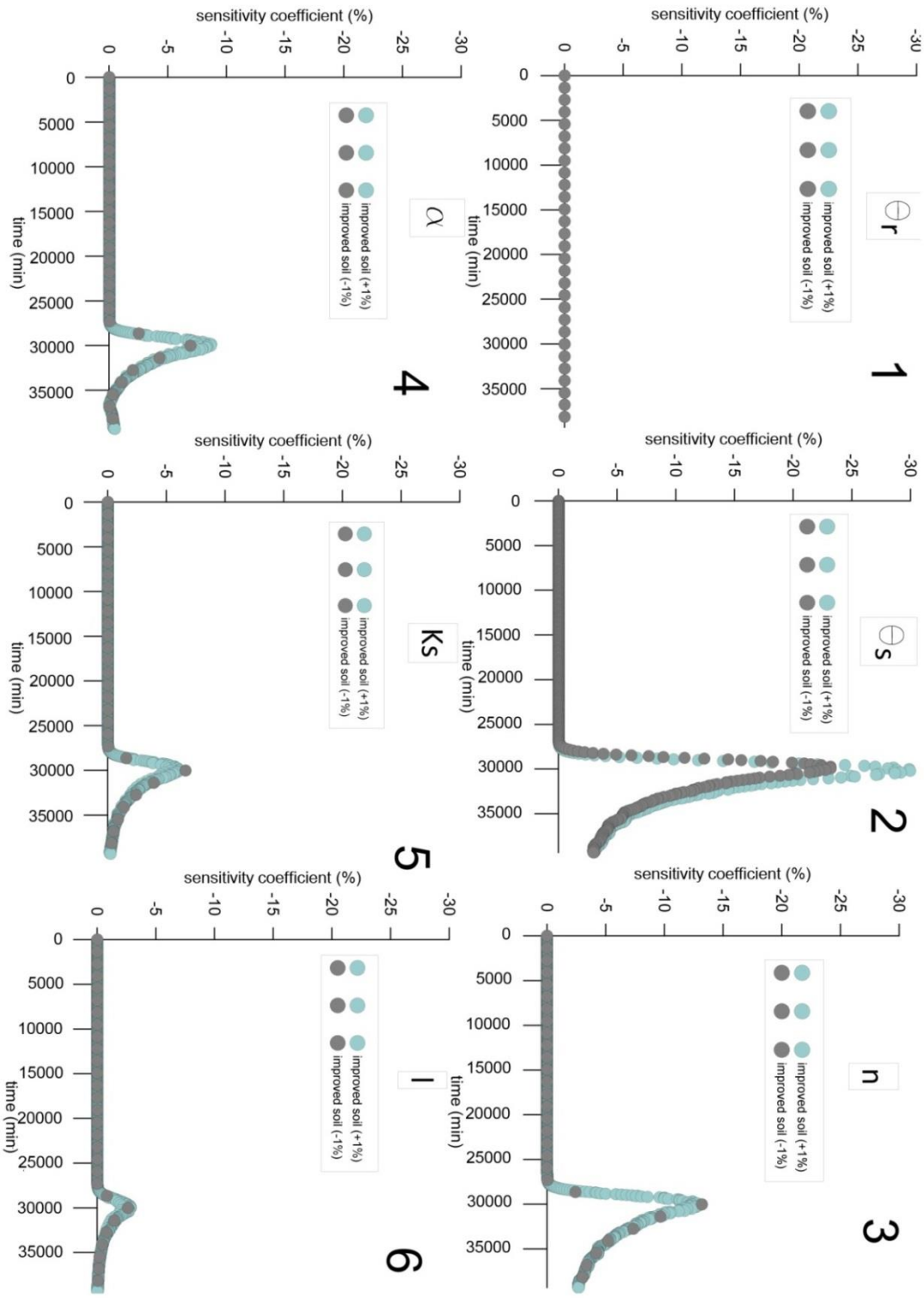


Table 68: Results of the OAT sensitivity analysis of the hydraulic parameters of the 1D filtration experiment: evolution of the sensitivity coefficient with time.

3_1_2_7 Calibration to replicate experimental results

30 infiltration simulations were performed to calibrate the hydraulic parameters of the mixed soil following the Levenberg-Marquardt (L-M) optimization algorithm implemented in Hydrus2D (Simunek et al, 2012b). For more information on the algorithm implemented in Hydrus2D to perform inverse modelling and the set of metrics/indices/properties selected to investigate indirect simulations performance see chapters §2_2 (PART II) and 2_7 (PART I) of the present thesis.

A summary of simulations run is presented in Table 69. As discussed by Russo et al. (1991), the inverse analysis has to be conducted several times with different initial parameters in order to verify that the problem is well posed and obtain a unique inverse solution. Consequently:

- simulations 1, 7, 13, 19 and 25 were conducted using the SWRC parameters obtained from the laboratory experiments presented in the present chapter §3_1_2. These parameters were refined in order to optimise the match between simulations and experimental observations.
- the rest of the simulations uses initial parameters taken from the USDA soil catalogue (Soil Science Division Staff, 2017). Again, the parameters were refined in order to optimise the match between simulations and experimental observations.

Table 69: Summary of inverse analyses performed to calibrate the 1D filtration model. In the second column the parameters optimized in each indirect simulation are reported; in the third column the type of observation points used (water pressure head with the acronym PH and water pressure head+water content with the acronym PH+wc); in the fourth column the weights given to the observation points; in the fifth column the initial parameters dataset and in the last column the optimized parameters (obtained from inverse). Note: PH: water Pressure Head, wc: water content.

Sim.n°	FITTED PARAMETERS	MEASURE-MENT SETS	WEIGHT	INITIAL ESTIMATES						FINAL ESTIMATES					
				θr	θs	α (1/m)	n	k (m/sec)	l	θr	θs	α (1/m)	n	k (m/sec)	l
1	θs; α; n; Ks; l	PH	all observed data w=1	0.00	0.45	4.26	1.42	1.98E-07	0.50	0.00	0.55	6.20	2.65	2.10E-07	0.25
2	θs; α; n; Ks; l	PH	all observed data w=1	0.07	0.41	7.50	1.89	1.23E-07	0.50	0.07	0.42	7.88	2.06	1.29E-07	0.51
3	θs; α; n; Ks; l	PH	all observed data w=1	0.08	0.43	3.60	1.56	2.89E-07	0.50	0.08	0.61	5.33	1.59	6.57E-07	0.50
4	θs; α; n; Ks; l	PH	all observed data w=1	0.10	0.39	5.90	1.48	3.64E-07	0.50	0.10	0.57	7.65	1.11	4.95E-07	0.27
5	θs; α; n; Ks; l	PH	all observed data w=1	0.09	0.43	1.00	1.23	1.94E-07	0.50	0.09	0.82	1.51	1.06	5.25E-07	0.30
6	θs; α; n; Ks; l	PH	all observed data w=1	0.03	0.46	1.60	1.37	6.94E-07	0.50	0.03	0.66	2.58	1.93	5.88E-07	0.83
7	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.00	0.45	4.26	1.42	1.98E-07	0.50	0.00	0.58	5.00	2.58	2.92E-07	0.50
8	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.07	0.41	7.50	1.89	1.23E-07	0.50	0.07	0.42	7.92	2.06	1.29E-07	0.50
9	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.08	0.43	3.60	1.56	2.89E-07	0.50	0.08	0.54	5.71	1.51	5.87E-07	0.50
10	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.10	0.39	5.90	1.48	3.64E-07	0.50						
11	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.09	0.43	1.00	1.23	1.94E-07	0.50	0.09	0.78	1.52	2.79	1.27E-07	0.50
12	θs; α; n; Ks	PH+wc	PH (w=1) wc (w=2)	0.03	0.46	1.60	1.37	6.94E-07	0.50	0.03	0.91	1.88	1.06	2.13E-07	0.50
13	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.00	0.45	4.26	1.42	1.98E-07	0.50	0.00	0.55	5.87	2.52	2.15E-07	0.25
14	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.07	0.41	7.50	1.89	1.23E-07	0.50	0.07	0.42	7.88	2.06	1.29E-07	0.51
15	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.08	0.43	3.60	1.56	2.89E-07	0.50	0.08	0.72	4.70	1.08	4.40E-07	0.02
16	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.10	0.39	5.90	1.48	3.64E-07	0.50	0.10	0.55	7.18	1.13	4.59E-07	0.28
17	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.09	0.43	1.00	1.23	1.94E-07	0.50	0.09	0.75	1.34	1.09	5.48E-07	0.33
18	θs; α; n; Ks; l	PH+wc	PH (w=1) wc (w=2)	0.03	0.46	1.60	1.37	6.94E-07	0.50	0.03	0.69	4.04	1.86	8.16E-07	0.54
19	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.00	0.45	4.26	1.42	1.98E-07	0.50	0.00	0.55	5.87	2.52	2.52E-07	0.25
20	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.07	0.41	7.50	1.89	1.23E-07	0.50	0.07	0.42	7.88	2.06	1.29E-07	0.51
21	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.08	0.43	3.60	1.56	2.89E-07	0.50	0.08	0.72	4.70	1.08	4.40E-07	0.02
22	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.10	0.39	5.90	1.48	3.64E-07	0.50	0.10	0.55	7.18	1.13	4.59E-07	0.28
23	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.09	0.43	1.00	1.23	1.94E-07	0.50	0.09	0.75	1.34	1.09	5.48E-07	0.33
24	θs; α; n; Ks; l	PH+wc	PH* (w=1) wc (w=2)	0.03	0.46	1.60	1.37	6.94E-07	0.50	0.03	0.69	4.04	1.86	8.16E-07	0.54
25	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.00	0.45	4.26	1.42	1.98E-07	0.50	0.00	0.53	3.61	2.17	1.98E-07	0.50
26	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.07	0.41	7.50	1.89	1.23E-07	0.50	0.07	0.41	7.78	2.08	1.23E-07	0.50
27	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.08	0.43	3.60	1.56	2.89E-07	0.50	0.08	0.56	4.41	1.13	2.89E-07	0.50
28	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.10	0.39	5.90	1.48	3.64E-07	0.50	0.10	0.57	5.48	1.12	3.64E-07	0.50
29	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.09	0.43	1.00	1.23	1.94E-07	0.50						
30	θs; α; n	PH+wc	PH* (w=1) wc (w=2)	0.03	0.46	1.60	1.37	6.94E-07	0.50	0.03	0.93	3.08	1.06	6.94E-07	0.50
* A weight equal to zero is set to all the observation points prior and after the passage of the moisture															

28 out of 30 simulations have been successful i.e. 28 simulations reached convergence in the inverse analysis.

- simulations 1 to 6 and 13 to 18 focused on optimizing parameters θs, α, n, Ks and l.
- simulations 7 to 12 and 13 to 18 focused on optimizing parameters θs, α, n, Ks .In the indirect simulations (7 to 12 ; 13 to 18) the pore connectivity factor ''l'' has been excluded due to a low sensitivity of the model to this parameter as observed in the sensitivity analysis.
- simulations 25 to 30 focused on optimizing parameters θs, α, n which are the hydraulic parameters that showed the highest sensitivity.

Two types of experimental observations were compared to the numerical output:

- simulations 1 to 6 relied on pressure head (PH) measurement at the first and second holes from the base of the column (9.5 cm and 15.5 cm). The weight associated to the PH observations is 1 for all data. The weight is the $w_{i,j}$ term in equation 44: the weight associated to each data point of the observed dataset.
- simulations 7 to 18 relied on pressure head (PH) measurement at the first and second holes from the base of the column and on the initial water content of the soil (6.5%). A weight of 2 was set to the initial soil water content.

- simulations 19 to 30 set a weight equal to zero to all the observation points prior and after the passage of the moisture. In fact, as observed in the sensitivity analysis after reaching saturation, the sensitivity of the model to change in the hydraulic parameters decreases rapidly to zero.

The calibration was conducted in two phases similarly to what has been was done for the Infrasafe project (see chapters §7_3; §7_4 and §7_5, PART I of the present thesis) but using simplifications due to the reduced number of simulations performed, the reduced number of observation points in the inverse problem and the greater simplicity of the model itself (one single material, low number of parameters to optimize, steady state 1D filtration). The objective of phase 1 is to evaluate the deviation between observed-simulated datasets by means of a set of metrics/indices applied for the evaluation of each indirect simulation performance. The chosen metrics are of typologies accuracy and efficiency. In the second phase, the focus is turned to the results of the inverse analysis and the obtained set of optimized parameters in the last iteration of the indirect problem and the accuracy with which the parameters have been estimated. Due to the fact that the aim of phase 2 is different from phase 1 the chosen parameters/coefficients are different.

The metrics used in calibration phase 1 in order to assess the goodness of fit (see chapter §2_7_2 of PART I for the full discussion on the indices/metrics) include:

- accuracy metrics as MAE (Mean Absolute Error), RMSE (Root Mean Square Error considering the different weighting distribution of the observed dataset), MSE (Mean Square Error), CRM (Coefficient of Residual Mass) and R^2 (the Coefficient of Determination).
- efficiency metrics as IA (the Index of Agreement)and its modifications (IA_j, IA'), NSE (Nash-Sutcliffe Efficiency) and its modified form NSE_j and the KGE (Kling-Gupta Efficiency) index.

The analytical expressions of the chosen metrics are reported in chapter §2_2 (PART II) of the present thesis.

The MAE, RMSE, MSE, CRM indices return a "zero" value for a perfect fit while the NSE, NSE_j, IA, IA_j, IA', KGE, R^2 metrics return a value of 1. To accommodate these differences in the assessment of goodness to fit, scores have been calculated for each inverse simulation: the best score is 1 and it is given to the simulation which shows a value closer to the best fit while the score of the other simulations is given proportionally dividing the value of the metric by the best value obtained among the 28 performed indirect simulations. For sake of clarity an example is reported below.

Let's consider the coefficient of determination R^2 calculated from the output of three performed indirect simulations A, B, C with the following values: 0.98; 0.95; 0.90. Simulation A shows the best value (R^2 closer to 1 that is the best fit) so its score is set to 1, while the score of simulation B is set to 0.95/0.98=0.969 and the one of the simulation C to 0.90/0.98=0.918.

Table 70 presents the different weights attributed to each metrics/index used in the 1st phase of the calibration process. The same weight (2.5) has been attributed globally to the set of accuracy and efficiency metrics while internally:

- The three groups of efficiency metrics (IA, IA_j, IA') and (NSE, NSE_j)and (KGE) have a weight of 0.833 each.
- The accuracy metrics have a weight of 0.5 each.

Weights have been set to each metric/index in order to ''balance'' the evaluation of the indirect simulation performance equally between the different typologies of metric (accuracy, efficiency) and internally between different formulations of the same metric (f.e. IA, IA_j and IA' have globally the same weight of NSE and NSE_j).

Table 70: Different weights attributed to each metric/index used in the 1st phase of the calibration

Accuracy metrics		Efficiency metrics	
MAE	0.50	**IA**	0.277
RMSE (considering the weighting distribution)	0.50	**IA_j**	0.277
MSE	0.50	**IA'**	0.277
CRM	0.50	**NSE**	0.416
R^2	0.50	**NSE_j**	0.416
Total weight accuracy metrics	2.5	**KGE**	0.833
		Total weight efficiency metrics	2.5
Total weight 1st phase calibration	5.0		

The final score s_{fin} of each indirect simulation at the end of the first phase of the calibration is computed as follows:

$s_{fin} = \sum_{k=1}^{n} w_k \frac{v_k}{v_{opt}} = \sum_{k=1}^{n} w_k \, s_k$ if the best fit of the considered kth metric is 1 [equation 52]

$s_{fin} = \sum_{k=1}^{n} w_k \frac{v_{opt}}{v_k} = \sum_{k=1}^{n} w_k \, s_k$ if the best fit of the considered kth metric is 0 [equation 53]

where n is the number of considered metrics/indices (11); $w_k$is the weight attributed to the k_{th} metric (Table 70); v_k is the value assumed by the k_{th} metric for the considered simulation, v_{opt} is the best value assumed by the kth metric among the pool of 28 indirect simulations, s_k is the score associated to the kth metric (score equal to 1 if $v_k = v_{opt}$). So the max score that could be attributed in the first phase to each simulation is 5. In Table 71 the values of some of the metrics of the 1st calibration phase are reported for each indirect simulation and in last six columns the score $w_k s_k$ calculated for each of the considered metrics.

In general the performance of the investigated inverse simulations is good, this is reflected in the values of the metrics of the first phase of the calibration (see Table 71): values of R^2 above 0.96 (best fit 1), IA values above 0.99 (best fit 1), NSE above 0.95 (best fit 1), NSE above 0.96 (best fit 1) are proofs of that. This means that the use of a PH head dataset obtained from 1D filtration is sufficient for a correct estimation of the hydraulic parameters.

Table 71: The table reports the values of some of the metrics/indices considered in the 1st phase of the calibration. The last six columns report the score $w_k s_k$ of each metric (see equations 52 and 53).

Sim.n°	R^2	CRM	MAE	KGE	IA'	NSE	SCORE R^2	SCORE CRM	SCORE MAE	SCORE KGE	SCORE IA'	SCORE NSE
1	0,98	-0,06	3,85	0,88	0,92	0,97	0,495	0,440	0,488	0,784	0,269	0,416
2	0,99	-0,09	3,82	0,85	0,92	0,96	0,497	0,269	0,492	0,755	0,270	0,414
3	0,99	-0,06	4,21	0,82	0,92	0,96	0,500	0,419	0,447	0,727	0,267	0,412
4	0,98	-0,11	3,76	0,84	0,92	0,96	0,496	0,225	0,500	0,746	0,270	0,414
5	0,99	-0,07	4,14	0,82	0,92	0,96	0,499	0,351	0,453	0,731	0,268	0,412
6	0,99	-0,06	4,21	0,82	0,92	0,96	0,500	0,419	0,447	0,727	0,267	0,412
7												
8	0,98	-0,09	3,80	0,85	0,92	0,96	0,496	0,277	0,495	0,762	0,270	0,415
9	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,422	0,448	0,728	0,267	0,412
10	0,96	-0,06	4,07	0,93	0,92	0,96	0,486	0,440	0,461	0,826	0,268	0,413
11	0,99	-0,07	4,08	0,82	0,92	0,96	0,499	0,344	0,460	0,733	0,268	0,412
12	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,422	0,448	0,728	0,267	0,412
13	0,98	-0,06	3,85	0,88	0,92	0,97	0,495	0,440	0,488	0,787	0,269	0,416
14	0,97	-0,05	3,87	0,91	0,92	0,97	0,491	0,500	0,485	0,809	0,269	0,415
15	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,419	0,448	0,728	0,267	0,412
16	0,98	-0,10	3,78	0,84	0,92	0,96	0,496	0,249	0,497	0,751	0,270	0,414
17	0,99	-0,07	4,08	0,82	0,92	0,96	0,499	0,344	0,460	0,733	0,268	0,412
18	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,419	0,448	0,728	0,267	0,412
19	0,99	-0,09	3,91	0,83	0,92	0,96	0,499	0,290	0,480	0,737	0,269	0,413
20	0,99	-0,06	4,20	0,81	0,92	0,96	0,500	0,408	0,447	0,726	0,267	0,412
21	0,98	-0,06	3,88	0,89	0,92	0,97	0,492	0,394	0,484	0,798	0,269	0,415
22	0,98	-0,06	3,85	0,89	0,92	0,97	0,494	0,434	0,488	0,792	0,269	0,416
23												
24	0,96	-0,05	4,26	0,93	0,91	0,95	0,482	0,497	0,441	0,833	0,267	0,411
25	0,98	-0,06	3,85	0,88	0,92	0,97	0,495	0,440	0,488	0,787	0,269	0,416
26	0,97	-0,05	3,87	0,91	0,92	0,97	0,491	0,500	0,485	0,809	0,269	0,415
27	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,419	0,448	0,728	0,267	0,412
28	0,99	-0,07	4,08	0,82	0,92	0,96	0,499	0,344	0,460	0,733	0,268	0,412
29	0,99	-0,07	4,08	0,82	0,92	0,96	0,499	0,344	0,460	0,733	0,268	0,412
30	0,99	-0,06	4,19	0,82	0,92	0,96	0,500	0,419	0,448	0,728	0,267	0,412
BEST FIT	1,00	0,00	0,00	1,00	1,00	1,00						

In Phase 2 the goodness of the indirect simulation performance was assessed in terms of:

- Sum of Squared Residuals (SSQ)

- Mass Balance Error (MBE)
- Standard Error Coefficient of the optimized parameters (SEC)
- number of iterations performed for each simulation (NoI)
- Confidence Interval of each optimized parameter (CI)
- Correlation Coefficient of each couple of optimized parameters (CC).

All these parameters are defined and discussed in great detail in chapter §7_4 (PART I) and in chapter §2_2 (PART II) of the present thesis.

The highest score possible for each parameter/coefficient of phase 2 is 1, as per simulations of phase 1. Then, the final score of phase 2 for each indirect simulation is calculated as in equations 52 and 53 using the following different weights for the chosen parameters/coefficients:

- SSQ: weight of 0.5;
- MBE: weight of 1;
- SEC: weight of 1;
- NoI performed for each simulation: weight of 0.5;
- CI of each optimized parameter: weight of 1;
- CC of the couple of optimized parameters: weight of 1.

Attributing weights to indices/metrics/indicators used to evaluate a model performance comes from the fact that some indices have an higher ability to describe the goodness of a simulation outcome in the process of parameter optimization. For instance, the NoI is not able to give information on the accuracy of the optimized parameters prediction but only it suggests how ''far'' the new set of optimized parameters is from the initial one. In other words NoI gives indication on the position of the optimized parameters with respect to the initial parameters in the space of the parameters. At the same way, SSQ is an index able to give information on the discrepancy between observed and simulated data, but these information are dealt with more accuracy in the first phase of the calibration and here the index is considered to give an overall-simplified indication of the outcome of phase 1. For all these reason a reduced weight (0.5) is set for NoI and SSQ.

For sake of clarity another example of how to compute the final score of an indirect simulation in the second phase of the calibration using equations 52 and 53 is reported for one of the chosen parameters/coefficients, the mass balance error. Let's assume to have performed three indirect simulations A, B, C and the mass balance error is equal to 0.5; 1; 1.5%, respectively. The performance that shows the best value is simulation A (mass balance error closer to 0). The score of the parameter MBE for each simulation is computed as follows:

$$s_{fin}(A) = w_k \frac{v_{opt}}{v_k} = 1\frac{0.5}{0.5} = 1 \qquad \text{[equation 54]}$$

$$s_{fin}(B) = w_k \frac{v_{opt}}{v_k} = 1\frac{0.5}{1} = 0.5 \qquad \text{[equation 55]}$$

$$s_{fin}(C) = w_k \frac{v_{opt}}{v_k} = 1\frac{0.5}{1.5} = 0.33 \qquad \text{[equation 56]}$$

Table 72: Scores for each metric/index/indicator of each considered inverse simulation. The second column is the score of the Sum of Squared Residuals (SSQ), the third column is the score of the Mass Balance Error (MBE); the fourth column is the score of the Standard Error Coefficient (SEC); the fifth column is the score of the Number of Iterations (NoI) and sixth column is the score of the Confidence Interval of the optimized parameters, the seventh column is the score of the Correlation Coefficient of the

couple of optimized parameters. The ninth, tenth and eleventh columns report the scores of the 1st and 2nd calibration phases and the last one the final score for each indirect simulation.

Sim.n°	score SSQ	score MBE	score SEC	score NoI	score CI	score CC	score ACCURACY METRICS (1st phase)	score EFFICIENCY METRICS (1st phase)	score (2nd phase)	Final score 1st+2nd PHASE
1	0.99	0.22	0.39	1.0	0.00	0.57	2.42	2.45	2.18	7.05
2	0.94	1.00	0.43	0.61	0.01	0.74	2.10	2.00	2.96	7.06
3	0.96	0.13	0.42	0.90	0.01	0.63	2.19	2.38	2.12	6.68
4	0.90	0.42	0.74	0.37	0.01	0.47	2.11	2.41	2.27	6.79
5	0.85	0.38	0.64	0.20	0.00	1.00	2.14	2.38	2.55	7.06
6	0.84	0.28	0.60	0.10	0.01	0.74	2.19	2.38	2.09	6.66
7										
8	0.95	0.38	0.35	0.56	0.01	0.67	2.21	2.43	2.16	6.79
9	0.86	0.52	0.36	0.39	0.01	0.56	2.20	2.38	2.08	6.65
10	0.87	0.26	0.20	0.54	0.01	0.40	2.25	2.48	1.58	6.30
11	0.84	0.38	0.53	0.20	0.00	0.81	2.15	2.39	2.24	6.77
12	0.84	0.28	0.45	0.10	0.01	0.54	2.20	2.38	1.74	6.32
13	0.92	0.43	1.00	0.37	0.01	0.62	2.15	2 39	2.69	7.23
14	0.97	0.23	0.34	0.56	0.00	0.56	2.44	2.47	1.90	6.81
15	0.89	0.23	0.39	0.54	0.00	0.59	2.20	2.38	1.93	6.50
16	0.92	0.43	0.63	0.37	0.01	0.62	2.15	2.41	2.33	6.89
17	0.86	0.40	0.59	0.20	0.01	0.81	2.15	2.39	2.33	6.86
18	0.84	0.28	0.60	0.10	0.01	0.74	2.20	2.38	2.09	6.66
19	0.88	0.42	0.19	0.10	1.00	0.40	2.14	2.40	2.50	7.03
20	0.84	0.28	0.35	0.07	0.69	0.35	2.18	2.38	2.13	6.69
21	0.97	0.30	0.33	0.68	0.48	0.36	2.33	2.46	2.31	7.10
22	0.99	0.28	0.28	0.66	0.56	0.37	2.40	2.46	2.32	7.17
23										
24	0.80	0.32	0.22	0.46	0.60	0.39	2.21	2 48	2.17	6.86
25	0.94	1.00	0.43	0.61	0.01	0.74	2.18	2.42	2.96	7.56
26	0.97	0.23	0.34	0.56	0.00	0.56	2.44	2 47	1.90	6.81
27	0.89	0.23	0.39	0.54	0.04	0.59	2.20	2 38	1.97	6.54
28	1.00	0.27	0.58	0.71	0.00	0.52	2.42	2.45	2.23	7.09
29	0.86	0.40	0.59	0.20	0.01	0.81	2.15	2 39	2.33	6.86
30	0.84	0.28	0.60	0.10	0.01	0.74	2.20	2 38	2.09	6.66
MAX SCORE							2.50	2.50	5.00	10.0

Looking at the final scores presented in Table 72: it can be seen that simulations 13 and 25 scored the highest.

Both simulations used initial parameters coming from the SWRC obtained in laboratory (see Figure 52) and the pressure head and initial water content as experimental observations. In accordance with van Genutchen and Simunek (1996) and Russo et al (1991), it has been observed also in this case that the best way to identify hydraulic parameters is using a mixed dataset of observation points in fact simulations 13 and 25 use both the available information (PH data + initial water content). To enhance the performance of the inverse modelling, a possible future evolution of the current equipment would be the installation of one or multiple moisture sensors in order to obtain directly water content measurements in time at multiple depths in the column.

Looking closely at optimized parameters of simulations 13 and 25, it is possible to observe an offset of the α parameter toward an higher value (lower air entry matric potential) for simulation 13 (5.87) and toward a lower value for the simulation 25 (3.61) with respect to the initial parameter dataset based on the laboratory tests (see Table 69). The n parameter shows for both the indirect simulations an offset toward higher values (2.52 for simulation 13 and 2.17 for simulation 25) and at the same way the parameter θs (0.55 for simulation 13 and 0.53 for simulation 25).

As seen in the calibration process related to the Infrasafe project, the importance of starting the calibration from an adequate set of hydraulic parameters is here corroborated. Even for a simpler model (1D infiltration column), selecting adequate initial parameters results in shorter calculation time and more efficient computations.

Moreover, it is relevant to use for the inverse analysis the observation data which are the most sensitive to the simulated phenomena and to parameters variation and selecting appropriately the weights given to the observation data. For example in the present case study, the higher score is shown by simulation 25 which assumed a weight equal to zero for all the PH dataset prior and after the passage of the moisture because the model is sensible to parameters variation only in the time interval between initial saturation and full saturation. The use of the initial water content (together with the PH observed dataset) allows to fix a point in the SWRC, improving in this way the identifiability of the optimized parameters. Including in the observed dataset insensible information will lead to similar model predictions obtained starting from different set of optimized parameters, causing problems of non-identifiability and non- uniqueness of the solution (van Genuchten and Simunek, 1996).

In Table 73 the optimized parameters of simulation 25 that showed the best score among the 30 performed ones (see Table 72) are presented. In the second row the percentages of variation of the optimized parameters with respect to the initial parameters are presented .The percentages of variation (PoV) are calculated as follows:

$$\mathrm{PoV} = \frac{v_{opt} - v_{init}}{v_{init}} * 100 \qquad \text{[equation 57]}$$

Where v_{opt} is the value of the optimized parameter and v_{init} is the value of the initial parameter.

Table 73: Optimized parameters of simulation 25. In the second row the percentage of variation PoV of each optimized hydraulic parameter with respect to the initial guess.

θr (-)	θs (-)	α (1/m)	n (-)	Ks (m/sec)	l (-)
0.00	0.53	3.61	2.17	$1.98*10^{-7}$	0.5
0%	22%	15%	53%	0%	0%

The VGM function describing the main wetting curve and the Mualem's hydraulic conductivity function of the investigated improved soil are presented in the following equations:

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1+|\alpha h|^n]^m} = \frac{0.53}{[1+|3.61h|^{2.17}]^{0.539}} \quad \text{h<0} \qquad \text{[equation 58]}$$

$$\theta(h) = \theta_s = 0.53 \qquad \text{h≥0} \qquad \text{[equation 59]}$$

$$m = 1 - \frac{1}{n} = 1 - \frac{1}{2.17} = 0.539 \qquad n > 1 \qquad \text{[equation 60]}$$

$$K_s(h) = K_s S_e^l [1 - (1 - S_e^{\frac{1}{m}})^m]^2 = (1.98 * 10^{-7}) S_e^{0.5} [1 - (1 - S_e^{1/0.539})^{0.539}]^2 \qquad \text{[equation 61]}$$

$$\mathrm{S_e} = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{\theta}{0.53} \qquad \text{[equation 62]}$$

Where S_e is the effective water content. In Figure 60 the observed data (from tensiometers readings) and simulated data (using the optimized parameters dataset of inverse simulation 25) in time are presented overlapped.

The hydraulic parameters calibrated using data of 1D filtration are the parameters of the main wetting curve because the sample is initially close to dry conditions then water is supplied until saturation is reached (process of imbibition) while the hydraulic parameters obtained from the laboratory technique presented in subchapter §3_1_2_2 describe the main drying curve because the sample is initially water saturated then dried (process of drainage).

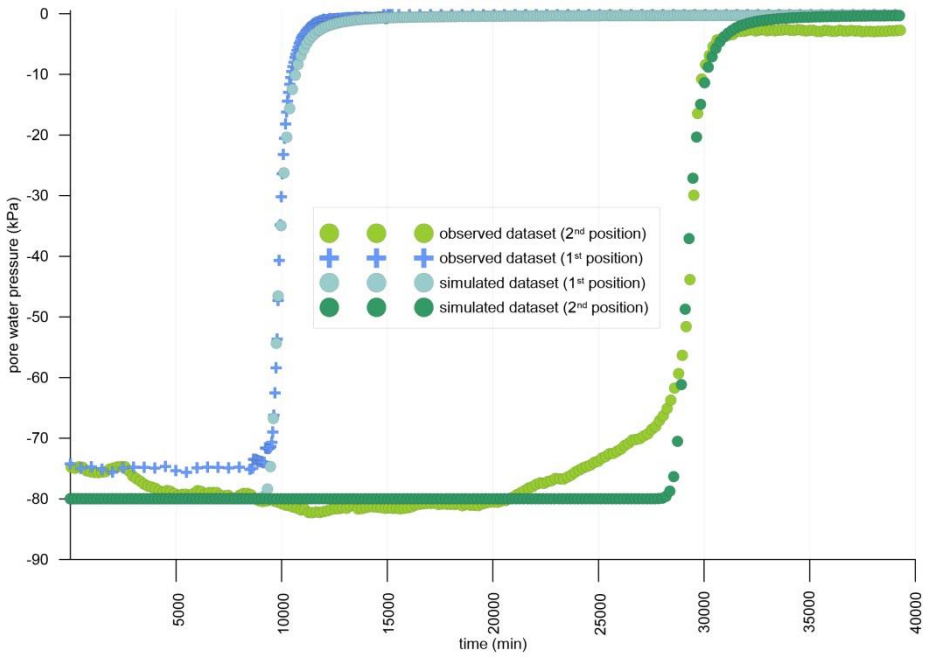


Figure 60: Observed and simulated data in the first and second observation point from the base of the filtration column. Simulation 25.

3_1_3 Potting Mix

3_1_3_1 Basic Geotechnical Properties

The potting mix in which the Melaleuca Styphelioides plant has been grown was also characterized as it is present in the large scale evapo-transpiration experiment. The potting mix is a blend of sphagnum peat moss, coir fiber, perlite, sand, limestone and fertilizers. The specific weight of the potting mix is equal to 1.679 (standard deviation 0.005). The organic content of the potting mix is equal to 54.47% (standard deviation of 10.22%). Atterberg limits do not apply to this sort of material (no fine particles). The material contains between 28.3% and 51.5% of particles of gravel size, between 42.6% and 66.6% of particles of sand size and between 5.1% and 5.8% of fine fraction. Table 75 summarizes the basic geotechnical properties of the potting mix. A summary of the main physical properties of the potting mix is presented in Table 74.

Table 74: Main geotechnical properties of the potting mix, organic soil in which the plants tested in laboratory have been grown.

	Average value	Standard deviation		Grain size distribution
Organic content	54.47	10.22	**% gravel**	28.34-51.52
G_s	1.679	0.005	**% sand**	42.63-66.58
Liquid limit (%)	/		**% fine particles**	5.08-5.85
Plastic limit (%)	/			
IP	/			

3_1_3_2 Soil water retention curve

The potting mix is cohesionless and contains large particles; it is hence very difficult to measure its shrinkage curve. Consequently, only its gravimetric retention curve was determined. This was achieved following the same procedure described for the improved soil: the potting mix was placed at a high void ratio (around 2) in a WP4 container and suction values higher than 1 MPa were measured using the dewpoint potentiometer WP4 while suction values below 1MPa were measured using high-capacity tensiometers. The obtained SWRC is at constant void ratio (~2) and it is representative of the drying branch of the retention curve because the wet initial material is let to dry increasingly to describe each point. So a new specimen has been prepared for each point of the retention curve, starting from the loose potting mix at a given water content.

Figure 61 shows the SWRC found for the potting mix. It is bimodal, which is thought to reflect the existence of microscopic and macroscopic porosities and is expected in such an highly-organic soil (Burger and Shackelford, 2001). The inflection point corresponds to the complete desaturation of the macroscopic porosity and the beginning of the desaturation of the microscopic porosity.

Dual porosity models assume that the porous medium consists of two distinct pore systems, both of them treated as homogeneous media with proper hydraulic properties. The two regions are considered as superimposed over the same volume (Dykhuizen, 1987). One region is associated with macro-pores (inter-aggregate or fracture of the rock matrix through which flow takes place), the other to micro-pores (or intra-aggregate pores) of the rock matrix. The liquid phase θ is divided in two contributes θ_{mo} the mobile one (inter-aggregate) and θ_{im} the immobile one (intra-aggregate) as follows:

$$\theta = \theta_{mo} + \theta_{im} \qquad \text{[equation 63]}$$

The water pockets of the intra-aggregate pores can exchange water with the inter-aggregates during drying process and retain water during wetting process but they could not exchange directly with the other intra-aggregate pores. For this reason water in the micro-pores is considered as immobile from a larger scale point of view (Simunek and van Genutchen, 2008).

The complexity of the bimodal retention curve lies in the necessity to fit the experimental data with an S-shaped curve such as the van-Genuchten model or Brooks-Corey model. To overcome this limitation, Durner (1994) proposed to fit data points before and after the inflection point suction value separately, leading to two sets of hydraulic parameters for the two regions (microscopic and macroscopic). The two superimposed VGM equations are weighted by their respective volume fractions of bulk soil ω_1 and ω_2 (Gerke and van Genuchten, 1993) as follows:

$$\theta(h) = \omega_1\theta_1(h) + \omega_2\theta_2(h) \quad \text{[equation 64]}$$

$$\theta_i(h) = \theta_{r,i} + \left(\theta_{s,i} - \theta_{r,i}\right)[1 + (\alpha_i h)^{n_i}]^{-m_i} \text{ with } i\text{=1, 2 and } m_i = 1 - \frac{1}{n_i} \quad \text{[equation 65]}$$

where the subscript i=1 stands for the macro-pores and i=2 for their micro-pores. $\theta_{r,i}, \theta_{s,i}, \alpha_i, n_i, m_i$ are the hydraulic parameters of the two separate functions.

$$S_e = \omega_1[1 + (\alpha_1 h)^{n_1}]^{-m_1} + \omega_2[1 + (\alpha_2 h)^{n_2}]^{-m_2} \quad \text{[equation 66]}$$

Where S_e is the effective water content of the superimposed regions.

Combining the Durner (1994) retention model with Mualem's hydraulic conductivity function the following equation is obtained:

$$K(S_e) = K_s \frac{(\omega_1 S_{e_1} + \omega_2 S_{e_2})^l (\omega_1\alpha_1[1-(1-S_{e_1}^{\frac{1}{m_1}})^{m_1}] + \omega_2\alpha_2[1-(1-S_{e_2}^{\frac{1}{m_2}})^{m_2}])^2}{(\omega_1\alpha_1 + \omega_2\alpha_2)^2} \quad \text{[equation 67]}$$

Bimodal SWRC models have been proposed to capture the contribution of both the porosities in the unsaturated zone (Kohne et al, 2009; Simunek et al,2003; Jarvis,2007) but, despite this, the unimodal VGM remains the most used model. The two sets of parameters of the potting mix for the microscopic and macroscopic regions considered separately are presented in the first and second line in Table 75 while in the third line the hydraulic parameters resulting from the simple application of the VGM to the experimental points without differentiation between the two porosity regions. In the last line the parameters of the dual-porosity model of Durner (1994). The free software RETC elaborated by M.Th. van Genutchen, J. Simunek , F.J.Leiji and M.Sejna (van Genuchten et al, 1991) has been used to estimate the soil water retention parameters of the potting mix for the different hydraulic models. RETC uses a nonlinear least-squares parameter optimization algorithm to predict the parameters of the hydraulic conductivity function and of the retention function starting from laboratory data.

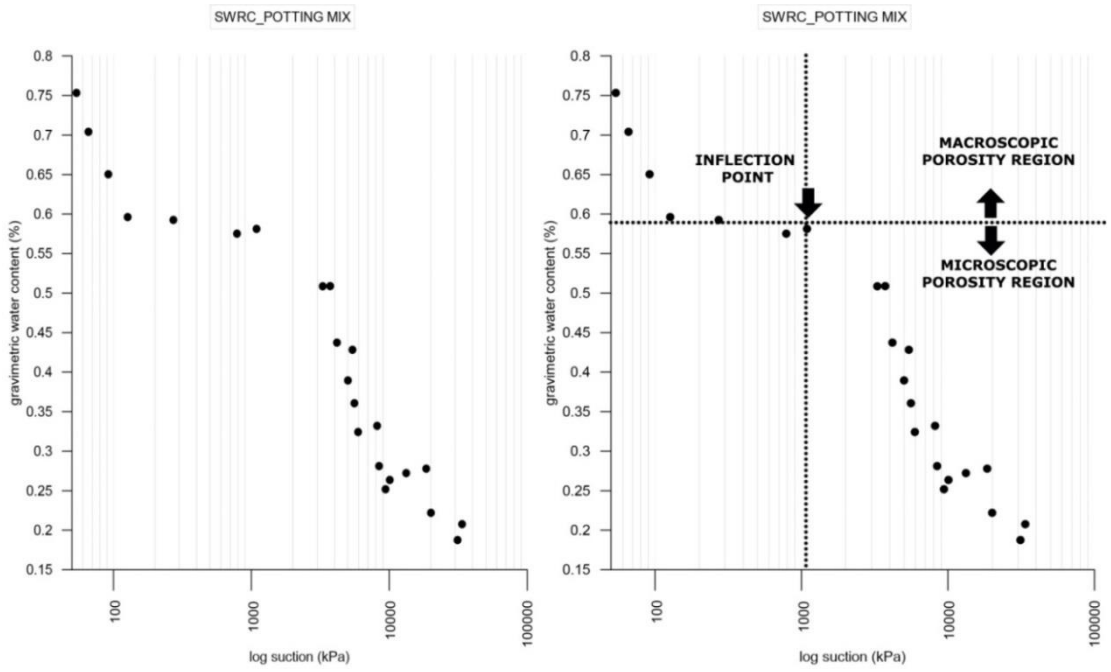


Figure 61: Bimodal SWRC of the potting mix (drying branch, void ratio ~2)

Table 75: Hydraulic parameters of the potting mix considering the unimodal VG model in the macroscopic porosity region (first line), in the microscopic porosity region (second line) and in both the regions (third line). In the fourth line the hydraulic parameters of the potting mix considering the dual porosity model of Durner, 1994. θr is the residual water content, θs the saturated water content, α and n water retention parameters (van Genuchten, 1980) of the inter-aggregate region, ω_2 is the relative weighting factor for the intra-aggregate region and α_2 n_2 two water retention parameters of the intra-aggregate domain.

	θ_r (-)	θ_s (-)	α (1/m)	n (-)	l (-)	ω_2	α_2 (1/m)	n_2
Macroscopic porosity region (unimodal VGM)	0.580	0.827	0.162	4.023	0.5	/	/	/
Microscopic porosity region (unimodal VGM)	0.188	0.779	0.004	2.493	0.5	/	/	/
Microscopic + Macroscopic regions (unimodal VGM)	0.000	0.661	0.005	1.453	0.5	/	/	/
Dual porosity model (Durner, 1994)	0.003	0.607	1.331	1.397	0.5	0.5	1.331	1.397

3_1_3_3 Saturated Permeability

The saturated permeability of the potting mix has been determined via a constant head permeability test, using the setup shown in Figure 62. The soil specimen has been created in the cylinder of plexiglass with a void ratio of 2, which corresponds to the void ratio of the potting mix in which the plants grew in the nursery. Three tests were carried out with the same sample of potting mix and using three different hydraulic gradients (see Figure 63). An average value of saturated permeability of $8.42 \cdot 10^{-6}$ m/sec was found. Using the Mualem (1976)'s capillary model, the hydraulic conductivity function is deduced indirectly from the SWRC function and the knowledge of the saturated permeability of the material.

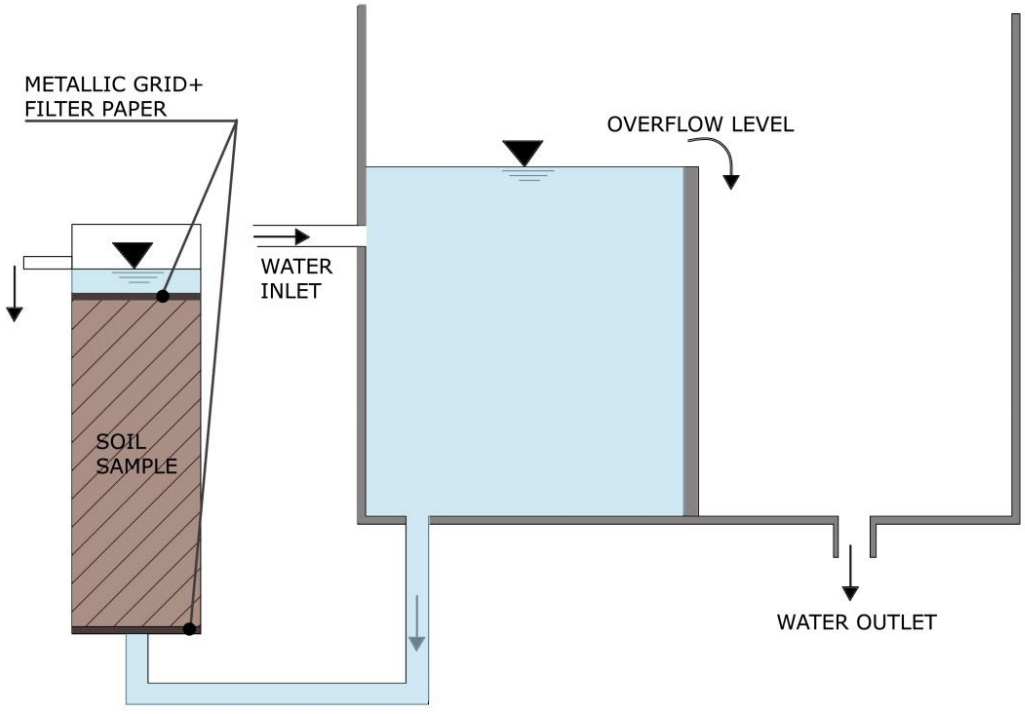


Figure 62: A scheme of the equipment to perform the constant head permeability test.

Figure 63: Two photos of the equipment during the testing.

3_1_4 Preliminary conclusions and further studies

Chapter 3_1 of PART II of the present thesis focuses on the physical and hydraulic characterization of the soil collected from the investigated site and of the improved soil used in the large scale apparatus together with the potting mix in which plant roots grow.

The investigated soil belonging to the Newcastle Inner City Bypass Cut 7, West Charlestown is an organic clay of high plasticity. The percentage of coarse material (sand) is equal to 26.5% and the percentage of fine material (clay+silt) is 73,5%.The saturated permeability of the soil is equal to 5.57E-08 m/sec. The soil is extremely acidic with a PH between 4,8 and 5,4 which could cause relevant problems to plants growth. The salinity of the soil is low as desirable. In order to ease the root growth, the original soil has been improved adding clean river sand in order to obtain a percentage of the coarse fraction equal to 65%. The higher quantity of macropores increases the availability of oxygen for plants. Moreover in order to fix the extreme acidity of the soil, lime was added to the mixed soil to the rate of the 0.04% of the dry mass in order to increase the pH up to ~ 6. The gravimetric retention curve of the improved soil has been investigated by means of laboratory tests. A saturated sample with a void ratio equal to the one in the large scale apparatus (0,8) has been prepared in the oedometer equipment then inserted in the plastic cup of the WP4. In steps measurements of suction by means of high capacity tensiometers are recorded as the sample dries (the sample dries then a period of time is waited to obtain an equalization of the moisture before measuring the suction). Below the cavitation pressure, suction measurements are carried out by means of the dewpoint method (WP4 by Decagon Device). The volumetric deformations upon drying have been investigated using the hand spry technique (Liu and Buzzi, 2014) in order to track the soil volumetric retention curve. The pore size distribution of the improved soil has been investigated using MIP. The structure is dominated by macropores, with an air-entry value of 2 kPa. The relatively large volume of macropores helps root growth and oxygen availability in the soil. The saturated permeability of the

improved soil has been determined in the Rowe Cell and it is equal to $2{,}93*10^{-7}$ m/sec. The unsaturated permeability function has been estimated by means of the transient profile method: experimentally a constant volume of water is fed from the base of a column of soil and the wetting upward front is monitored by means of high-capacity tensiometers. A numerical model reproducing the pwp and water content changes in time has been elaborated by means of Hydrus 2D. The parameters of the SWRC and of the permeability function are progressively tuned until the numerical results match the experimental observations. The numerically calibrated permeability function will then be taken as the permeability function of the soil. The experiment carried out investigates the wetting branch of the SWRC. Before tuning the parameters, a sensitivity analysis has been performed. One at a time, hydraulic parameters have been varied of +/-1% and the sensitivity coefficients calculated (OFAT analysis). It has been observed a very high sensitivity of the parameters θs and n while l and θr show a sensitivity close to zero. Moreover the parameters sensitivity is zero before (soil at the initial water content) and after (full saturation) of the hydraulic stimulus (upward gradient). Inverse analysis have been performed on the hydraulic parameters that showed high sensitivity. 30 indirect simulations have been run and 28 of them reached convergence. One part of the indirect simulations has been performed starting from the set of initial parameters individuated by laboratory test (SWRC), the other part assumed the initial guess of the parameters from the USDA soil catalogue. Two different observed datasets have been used: suction readings in time and suction + water content readings (mixed dataset). Different weighting distributions have been adopted: all weights equal to unity, double weight for the wc data and equal to unity for PH readings in the mixed dataset, weight equal to zero in the period before and after the transition of the hydraulic stimulus. Similarly to what has been done in PART I of the present thesis, the performance of inverse analysis has been investigated in two phases: the first phase looks at the differences between observed-simulated datasets by means of a series of indices/metrics (accuracy and efficiency metrics) while the second phase looks at the results of the inverse analysis and at the optimized parameters. It has been observed that the best performance belongs to the mixed dataset (in accordance with van Genutchen and Simunek (1996) and Russo et al (1991)) in fact the use of the initial water content (together with the PH observed dataset) allows to fix a point in the SWRC, improving in this way the identifiability of the optimized parameters. The simulation with the best performance has been obtained with a null weight of the data before the arrival of the hydraulic stimulus and after the complete saturation, in accordance to what was found in the sensitivity analysis. Adding insensible data (as the data before and after the hydraulic stimulus) leads to problems of non-identifiability and non- uniqueness of the solution, giving no improvement to the inverse analysis. Moreover it has been observed, similarly to what was found in PART I of the present thesis, that a reliable set of initial parameters from proper laboratory tests assures the best performance in the indirect problem with better results with respect to the use of an initial guess of the parameters with approximated methods (f.e. pedo-transfer functions). The potting mix in which the plant of the large-scale apparatus has been grown has been investigated from a physical and hydraulic point of view. The potting mix is a blend of sphagnum peat moss, coir fiber, perlite, sand, limestone and fertilizers and its organic content is equal to 54,47%. Similarly to what has been done for the improved soil, the gravimetric retention curve of the material has been determined. Sample were created at a void ratio of 2 (the void ratio of the potting mix in the large-scale apparatus) and starting from a water content close to saturation, the material is let to dry per step and contextually measurements of suction by means of high-capacity tensiometers are performed. Below cavitation pressure, dewpoint method has been used to determine the dry end of the SWRC (up to 1MPa).The investigation of the volumetric deformations of the material upon drying could not be executed due to its coesionless nature (impossibility to investigate the volumetric soil water retention curve). The obtained gravimetric soil water retention curve of the potting mix is bimodal reflecting the existence of

microscopic and macroscopic porosities as expected in such an highly-organic soil. The complexity of the bimodal retention curve lies in the necessity to fit the experimental data with an S-shaped curve such as the van-Genuchten model or Brooks-Corey model. To overcome this limitation, Durner (1994) proposed to fit data points before and after the inflection point in the SWRC separately, leading to two sets of hydraulic parameters for the two regions (microscopic and macroscopic). The free program RETC which uses a nonlinear least-squares parameter optimization algorithm to predict the parameters of the HCF and of the SWRC starting from laboratory data, has been applied to the laboratory SWRC using Durner and VG models. The saturated permeability of the potting mix has been determined by means of constant head permeability test and it is equal to $8{,}42 \cdot 10^{-6}$ m/sec.

In the following chapter 3_2, the numerical model of the large scale experiment elaborated using the commercial software Hydrus 2D has been presented. Details on the assumed geometry, boundary conditions, initial conditions, adopted RWU and stress reduction function parameters, the chosen hydraulic parameters, space discretization and iteration criteria are presented extensively.

3_2 NUMERICAL MODEL USING HYDRUS 2D

3_2_1 Geometry and boundary conditions

A 2D section containing the root bulb and the control zone has been modelled to compare numerical simulations to experimental data for the different boundary condition tested (see Figure 64). The section is divided in three regions: the layer of gravel on the bottom of the container, the main part of the box occupied by the improved soil with the root bulb in the centre and the control zone on the right hand-side. Consistent with the experiment, an impervious material separates the control zone from the zone in which the plant lives to avoid any influence of the roots on the pore pressure / water content distribution in the control zone. The root bulb and the potting mix have been modelled, as per experimental conditions. The root bulb has the same dimension of the pot in which the plant has been grown in the nursery. Observation points are taken at the exact location of sensors in the box.

A variable pressure head was applied at the bottom of the box (to represent the water table) and an atmospheric boundary condition was applied to the soil surface, as per experimental arrangement. The sides of the models are no flux boundary conditions (no exchange of liquid or vapour phase).

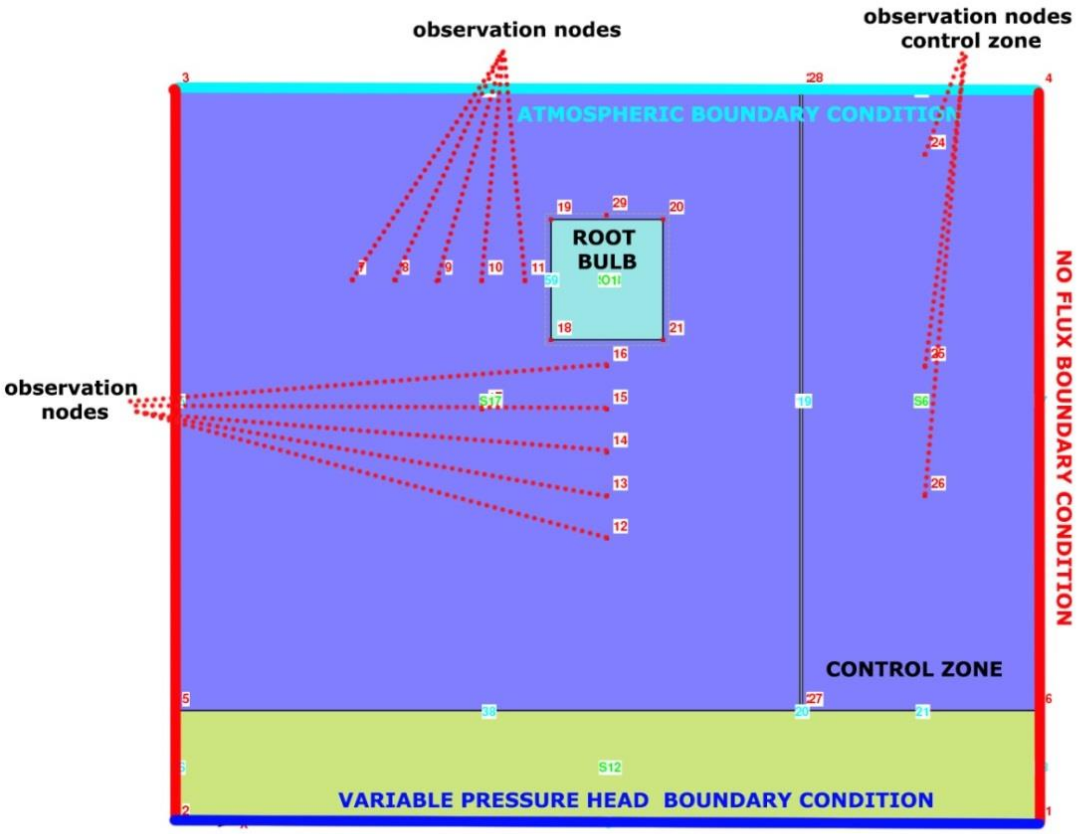


Figure 64: Schematic view of the large-scale apparatus with indications of the position of the observation points, of the boundary conditions and the different layers.

3_2_2 Evaporation and plant transpiration contributions

The evaporation and transpiration contributions were computed using the Penman-Monteith method (1948) in the last revision by Allen et al (1994) which expresses the Potential Daily Reference Evapo-transpiration (ET_o) as a function of the environmental conditions (radiation, air temperature, air humidity, wind speed). The ET_o is calculated using the free source software provided by FAO, ETo Calculator (Food and Agriculture Organization of the United Nations, 2009). Hydrus 2D needs the subdivision of the potential evapotranspiration contribution (ET_o) into the potential evaporation (E_{pot}) and potential transpiration (T_{pot}) contributions. Then the actual evaporation is computed by the code and it depends on the moisture state of the soil, becoming maximum after irrigation and decreasing as the surface dries due to a reduction of the permeability according to the hydraulic conductivity function of the soil. E_{pot} and T_{pot} are calculated using the following equations (FAO, 1996)

$$T_{pot} = ET_0[1 - \exp(-k\ LAI)] = ET_0 SCF \qquad \text{[equation 68]}$$

$$E_{pot} = ET_0 \exp(-k\ LAI) = ET_0(1 - SCF) \qquad \text{[equation 69]}$$

Where LAI is the leaf area index (dimensionless); SCF is the soil cover fraction; k is the radiation extinction coefficient by the canopy that usually is assumed as 0.463 (Ritchie, 1972) and ET_0 is the reference evapo-transpiration. SCF is the fraction of the soil within the unit area covered by vegetation canopy. LAI is the measure of the total area of leaves per unit ground area. LAI (m^2/m^2) ranges from 0 (bare ground) to 10 (dense forest).

Two evapo-transpiration contributions ($\boldsymbol{ET_o}$) have been computed. $\boldsymbol{ET_o}$ for the potential evaporation contribution is calculated using the environmental conditions above the soil surface. The radiation and the wind speed are equal to zero due to the fact that the upper side of the box is covered and sealed in order to maintain a constant RH (relative air humidity). The input climatic data are the RH and temperature measured by the two sensors installed in the enclosed space (see section on the description of the apparatus 3_5).

The potential evaporation rate is equal to the actual evaporation rate when the (negative) pressure head at the atmospheric boundary is greater than the minimum allowed (negative) pressure head value (hCritA). For values of (negative) pressure head lower than hCritA, the evaporation contribution is decreased from the potential value because the soil permeability has decreased to the point where the potential evaporation rate cannot be delivered anymore. hCritA is calculated from meteorological variables such as relative humidity according to the following expression (Feddes et al, 1974):

$$h_{CritA} = \frac{e^{RH}_{R\ T}}{M\ g} \qquad \text{[equation 70]}$$

Where RH is the relative air humidity (-); M is the molecular weight of water (0.018015 kg/mol), g is the gravitational acceleration (9.81 m/sec^2); R is the universal gas constant (8.314 J/(mol*K)), T is the absolute temperature (K).

For an average RH of 75% and a soil temperature of 16°C , h_{CritA} is equal to -2.8*10^4 m. In sandy soils, a small change in water content in the dry range leads to large changes in (negative) pressure head and in this condition maintaining the same boundary flux equal to the potential evaporation rate could cause instability in the code, for this reason it is always advisable to choose a h_{CritA} correspondent to a water content at least 0.005 higher than the residual water content (Simunek et al, 2012b).

To overcome this issue, a value of -8.4*10^3 m was assumed for the improved soil, as per recommendation by Simunek et al (2012b).

$\boldsymbol{ET_o}$ for the potential transpiration contribution is calculated using the input meteorological conditions to which the leaves of the plant are subjected (in this case, temperature and relative humidity of the room in which the box is stored and radiation from the agricultural lamp in use).

Because the experiment has been conducted over a relatively short duration and no significant plant growth has been observed, the coefficient LAI (and SCF) has been assumed to be constant through the tests. LAI has been measured by direct method dividing the area of the vegetated fraction (0,013 m^2) by the soil surface unit (1m^2) which is the area of the large scale apparatus.

3_2_3 Root water uptake spatial distribution and stress reduction function

The macroscopic root water uptake scheme proposed by Feddes et al (1978) in which the potential transpiration is distributed over the root zone and reduced to the actual root water uptake according to a stress reduction function α(h) has been adopted in Hydrus 2D (see chapter §2_1_2_5). The distribution of the root water uptake is based on the root distribution and water availability.

The parameters of the Feddes' α(h) function for the present numerical model can be found in Table 76. See chapter §2_1_2_5 for more information on α(h) function and see Figure 65 for an explanatory image of the

Feddes'parameters. The Feddes' parameters have been chosen from the Hydrus internal database based on studies of Wesseling (1991) and Taylor and Ashcroft (1972).

Table 76: Summary of the parameters of the water uptake reduction function α(h) of Feddes et al (1978) for the investigated case study.

h_1 [m]	h_2 [m]	h_{3high} [m]	h_{3low}[m]	h_4 [m]	T_{p1} [m/day]	T_{p2} [m/days]
-0,1	-0,25	-4	-4	-160	0,005	0,001

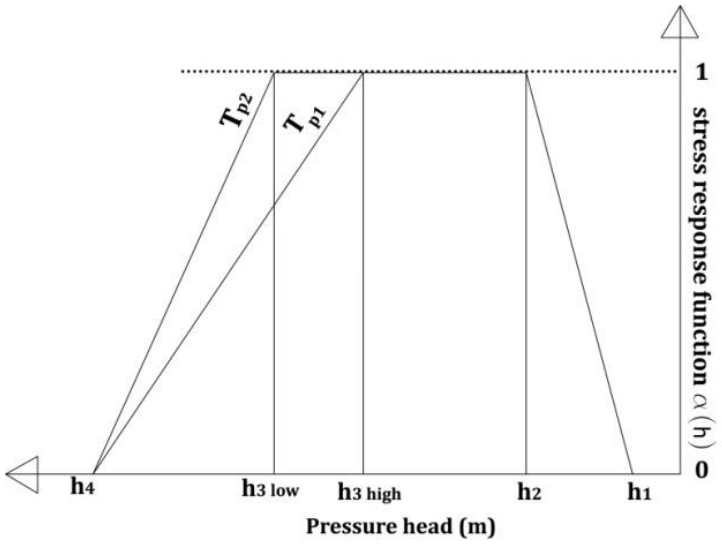


Figure 65: Graphical representation of the stress response function of Feddes et al (1978) with indication of all its parameters.

Where h_1 is the value of pressure head above which roots start to extract water from the surrounding soil; h_2 is the value of pressure head below which roots start to extract water at the maximum possible rate, h_{3high} is the value of pressure head below which roots can no longer extract water at the maximum rate (the potential transpiration rate is r_{2H}), h_{3low} is the value of pressure head below which roots can no longer extract water at the maximum rate (the potential transpiration rate is r_{2L}), h_4 is the wilting point. In Feddes' model h_3 is dependent on the potential transpiration rate (r_{2H} and r_{2L}). Indeed, as observed in nature, h_3 decreases at higher transpiration rate.

The 2D root spatial distribution function $b(x,z)$ of Vrugt et al (2001) implemented in Hydrus2D has been used to define the spatial variation of RWU in the root zone. The main function $b(x,z)$ is proposed again for a sake of clarity but more information could be found in chapter §2_1_2_2 (PART II).

$$b(x,z) = (1-\frac{z}{Z_m})(1-\frac{x}{X_m})e^{(\frac{p_z}{Z_m}|z^*-z|+\frac{p_r}{X_m}|x^*-x|)} \quad \text{[equation 71]}$$

where Z_m is the maximum rooting depth [L], X_m is the maximum rooting length [L] in the horizontal direction; p_z and p_r are empirical parameters (-) ; z^* and $x^*$are other empirical parameters [L]. Parameters p_z and p_r are set to unity for values of $z > z^*$ and $x > x^*$, respectively. $b(x,z)$ is computed in each point of the domain affected by a RWU and it is dimensionless.

The root water uptake decreases with increasing distance from the main root bulb to reflect the diminishing root density away from the bulb. The actual root water uptake is assumed to be dependent not

only on root distribution but also on the soil water availability. As the root zone (soil volume permeated by roots) is depleted of water, the difference between potential RWU and actual RWU increases.

One inherent difficulty of RWU models lies in the measurement of its parameters, which is non-trivial, to say the least. Most parameters are empirical and do not necessarily reflect the real properties or characteristics of the root system. For this reason, model parameters are usually adjusted until a good agreement is found between observations and simulated data (Cai et al, 2017) and this is the approach that was followed here.

Table 77 summarizes the spatial parameters adopted prior to the proper calibration that will use experimental data from the large scale apparatus (see chapter §3_3_1). The spatial distribution function $b(x,z)$ of Vrugt et al (2001) using the parameters reported in Table 77 is presented in Figure 66.

Table 77: Summary of the parameters of the RWU spatial distribution function of Vrugt et al (2001) for the investigated case study.

Parameters for vertical distribution		
Maximum rooting depth Z_m	**Depth of maximum intensity z^***	**Parameter p_z**
0.4	0.22	2.4
Parameters for horizontal (radial) distribution		
Maximum rooting radius X_m	**Radius of maximum intensity x^***	**Parameter p_r**
0.7	0.5	13

The parameters of the spatial distribution of RWU have been set according to the initial dimension of the root bulb. However, it is relevant to recall that the seedling has been planted in the soil about one month before the first controlled experiment and corresponding simulation. Roots respond quickly to external stress/stimuli (scarcity or abundance of water content or nutrients or oxygen) using sensory mechanisms to detect the osmotic status of the soil (Robbins et al, 2015) which leads to changes in the root bulb. The elongation of the fine roots, which depends on species and season, ranges from few millimetres to few centimetres per day (Nodichao et al, 2011).

During the initial period an horizontal development of new fine roots is probable due to an irrigation from the top along the stem which produced a sphere of wet soil around the bulb after each irrigation. Subsequently, the plant has been subjected to natural sub-irrigation (moisture is provided by the capillary rise from the water table), which prompts a downward root growth. The actual extent of the root bulb will be checked once the experiment has been completed.

The Hydrus 2D code focuses only on the water and solute transport and it does not take into consideration the crop and root growth and the various growth stages of a plant. The proliferation of fine roots is quite a fast process and in a long term experiment, it is necessary to periodically recalibrate the spatial distribution of RWU to reflect the possible growth of the root bulb. In the modelling of the large-scale apparatus experiment the root water uptake spatial distribution parameters have been calibrated using the available water potential measurements from the installed sensors. Due to the fact that the experiment is of relatively short duration a recalibration of the RWU spatial distribution was not necessary.

The root water uptake spatial distribution can be considered symmetrical with respect to the stem of the plant because the plant is subjected to symmetrical stimuli.

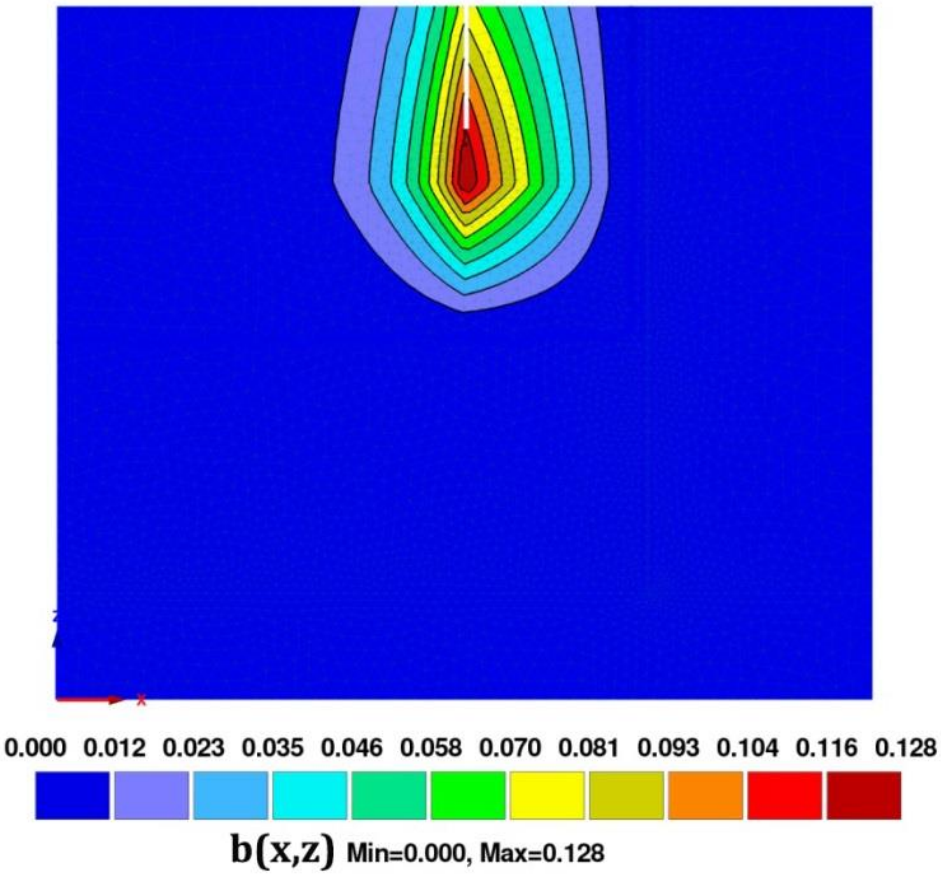


Figure 66: The spatial distribution function $b(x,z)$ of Vrugt et al (2001) using the parameters reported in Table 77. The spatial distribution function ranges from 0.128 in the centre of the root bulb to 0 (no influence of plant transpiration)

3_2_4 The hydraulic models (SWRC and hydraulic conductivity function)

The van Genuchten hydraulic model (VGM) has been chosen to simulate the moisture movement in the soil. If in one hand this choice implies to neglect soil hysteretic effects, on the other hand this simplification allows to deal with a lower number of model parameters. Globally the model requires the definition of 18 parameters: 5 parameters to define the SWRC and the hydraulic conductivity function, 6 parameters to define the root spatial distribution function of Vrugt et al (2001) and 7 parameters to define the water stress reduction function of Feddes et al (1978).

The present numerical model adopts as hydraulic parameters the ones of the main drying curve estimated from the laboratory data of soil water content vs water potential obtained for the investigated soil mix at the same void ratio of the soil in the box (0.8) and the saturated permeability obtained in the Rowe Cell (see chapters §3_1_2_2 and §3_1_2_3). For each material (improved soil and potting mix), six parameters have been determined: the saturated permeability Ks, the saturated water content θ_s, the residual water content θ_r, the tortuosity parameter of the conductivity function I, and finally, n and α, which are two empirical coefficients affecting the shape of the soil water retention curve. The hydraulic parameters have been estimated by the least square method fitting the VG equation, using the commercial RETC Code (van Genuchten et al, 1991) as presented in chapter §3_1_2_2.

For the gravel placed under the improved soil, pedotransfer functions PTFs (implemented in the computer program Rosetta[5]) were used to translate basic soil data (specific gravity and grain size distribution) in hydraulic properties used in the model. It is important to highlight that the only relevant hydraulic parameter for the gravel is the saturated hydraulic conductivity because for the full duration of the experiment the material remains in full saturation (below the water table working as a water reservoir).

Due to the fact that the assumed hydraulic model is the unimodal VGM, the dual porosity of the potting mix could not be considered because the same hydraulic model has to be adopted for all the materials of the domain. The hydraulic parameters of the SWRC have been obtained fitting together the data suction-water content of the macroscopic region and of the microscopic region and the resulting parameters are presented in the third line of Table 75.

A summary of the sets of initial hydraulic parameters of each layer, before the calibration programme performed in chapter §3_3_1 using the experimental data from the large-scale apparatus, is presented in Table 78.

Table 78: Sets of initial hydraulic parameters of the improved soil, potting mix and gravel, before the performance of the calibration programme using experimental data from the large-scale apparatus.

	θr [-]	θs [-]	α [1/m]	n [-]	Ks [m/sec]	l [-]
Improved soil	0.0001	0.45	4.26	1.422	$2.93*10^{-7}$	0.5
Potting mix	0.0000	0.66	0.0052	1.453	$8.42*10^{-6}$	0.5
Gravel	0.0650	0.41	8.00	1.890	$1.22*10^{-5}$	0.5

The parameters reported in Table 78 have to be substituted in the following equations of the van Genuchten model and Mualem (1976)'s model to define the SWRC and the hydraulic conductivity function, respectively. By way of example the parameters of the improved soil are substituted in the equations.

$$\theta(h) = \theta_r + \frac{\theta_s - \theta_r}{[1+|\alpha h|^n]^m} = \frac{0.45-0.0001}{[1+|3.61h|^{1.422}]^{0.296}} \quad h<0 \qquad \text{[equation 72]}$$

$$\theta(h) = \theta_s = 0.45 \quad h\geq 0 \qquad \text{[equation 73]}$$

$$m = 1 - \frac{1}{n} = 1 - \frac{1}{1.422} = 0.296 \qquad n > 1 \qquad \text{[equation 74]}$$

$$K_s\,(h) = K_s S_e^l [1 - (1 - S_e^{\frac{1}{m}})^m]^2 = (1.98 * 10^{-7}) S_e^{0.5} [1 - (1 - S_e^{1/0.296})^{0.296}]^2 \qquad \text{[equation 75]}$$

$$\mathrm{S_e} = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{\theta - 0.0001}{0.45 - 0.0001} \qquad \text{[equation 76]}$$

In chapter §3_1_2_7 the 1D filtration experiment has been used to calibrate the hydraulic parameters of the main wetting curve and of the hydraulic conductivity function, by means of inverse analysis. The indirect simulation 25, which showed the best performance, optimizes α, θ_s, n (of the main wetting curve) while the saturated permeability remains unchanged and equal to the value obtained experimentally in a Rowe Cell (chapter §3_1_2_3). These parameters characterizing the main wetting curve could be adopted

[5] Rosetta is a computer program to estimate soil hydraulic parameters using 5 hierarchical pedotransfer functions which require an increasing number of information for the estimation of the VGM retention parameters (Schaap et al, 2001)

in a possible future development of the model in which hysteretic effects and RWU effects are taken into account simultaneously. The present model deals solely with plant evapo-transpiration effects.

The simulation of the moisture movement starts with phase 1A in which the water table has been applied at the base of the large scale apparatus (water level of 0.2 m from the base of the apparatus) and it ends with the flooding of the plant and the sensors (phase 4).

Moisture movement was simulated for phases 1 to 3 to refine the calibration of the parameters of the Vrugt's model (for RWU) and van Genuchten model (improved soil, potting mix). The calibration phase has the main purpose of optimizing the agreement between observed and simulated dataset by means of inverse parameters estimation using the Levenberg-Marquardt algorithm. 23 indirect simulations were performed for the optimization of the hydraulic parameters of the improved soil: 6 simulations to optimize one single parameter; 9 simulations to optimize two parameters at a time; 7 simulations to optimize three parameters at a time; 1 simulation to optimize all parameters of the improved soil.

The same simulations were conducted to optimize the parameters of the potting mix.

Lastly, the hydraulic parameters of the improved soil and the potting mix were optimized together using the same combination of parameters (a single parameter of each material, two parameters of each material, three parameters at a time of each material and all the parameters of the two materials). The observed dataset used for the inverse simulations consists of the 14 observation points available and it is the same for all the inverse simulations carried out. The overall number of simulations is 69, of which 38 achieved convergence. This high number of successful simulations is an indication that the elaborated 2D model of the laboratory experiment is simpler than the model of the river bank elaborated for the Infrasafe project resulting in quicker convergence. Further details and the results of the inverse analysis performed on the hydraulic parameters of the 2 materials are exposed extensively in chapter §3_3_1_5.

3_2_5 Initial conditions

Initial conditions are recreated starting from the available data of suction in the observation points interpolating them and extrapolating pore pressure values in the unknown points. This allows to have an almost perfect match between observed-simulated data in the observation points at the initial time instant of the simulation. The starting instant of the simulation coincides with the beginning of phase 1A: the last irrigation of the plant along the stem has been performed, one instant before the application of the water level to the bottom boundary of the large-scale apparatus. Information on the performed phases of the experiment are presented in chapter §3_6.

3_2_6 Space discretization, time discretization and iteration criteria

An unstructured mesh containing 8180 triangular elements (dimension ranging from 0.003 to 0.04 m) and 4162 nodes has been created for the whole geometry (see Figure 67). The mesh is finer in and around the root bulb, around the water table, close to the soil surface and in vicinity of the sensors. These are zones where the model could give convergence problems if a very fine mesh is not introduced.

The time discretization associated with the numerical solution is described using three temporal parameters:

- the initial time step Δt_{init}, set at 0.9s. It is the initial time increment of the calculation, used at the beginning of the calculation and every time a sudden and significant change in the boundary conditions is detected.
- the minimum allowed time step Δt_{min}, set at 0.025s. This is the lower bound of all time increments in the computation.
- the maximum allowed time step Δt_{max}, set at 0.25 hour. This is the upper bound of all time increments in the computation.

Iteration criteria are used in Hydrus 2D/3D to iteratively solve Richards' equation and find the solutions of the global matrix equation at each time step. The iterative process stops when a satisfactory degree of convergence is obtained, i.e. the absolute change in pressure head (or water content) in each point of the domain between two successive iterations is less than the imposed tolerance (in terms of pressure head or water content).

The user can specify:

- a maximum number of iterations allowed for each time step, here 1000.

- a tolerance on water content and pressure head: 0.01m for pressure head and 0.05 (-) for water content.

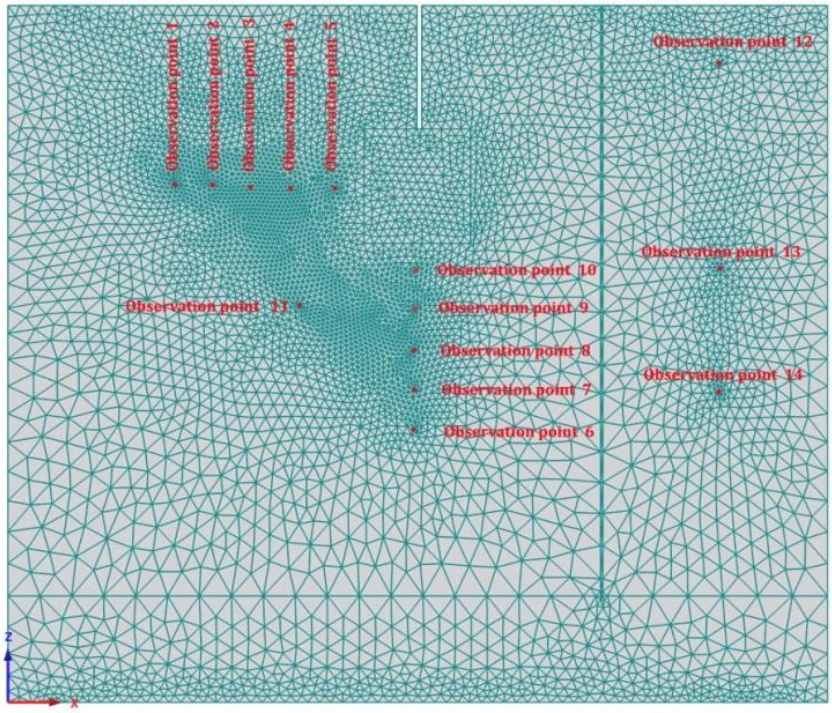


Figure 67 View of the mesh used to discretise the 2D medium in which moisture movement is simulated.

3_2_7 Preliminary conclusions and further studies

The experiment of the large scale apparatus has been reproduced numerically by means of the commercial code Hydrus2D. The geometry of the model is composed of a volume of improved soil in which, in the centre, the potting mix (with the dimension of the root bulb) has been set. An impervious surface divides the model in two parts: one part which contains the root bulb and the other part represents the control

zone and it is isolated from the roots influence. At the base of the apparatus a layer of gravel which distributes uniformly the water table in each point of the domain is set. The upper boundary is an atmospheric boundary condition subjected to the evaporation phenomenon, the boundary on the bottom of the large scale apparatus is a ''variable head'' boundary condition and the different hydraulic heads are here imposed. The conditions on the sides of the box are ''no-flux'' boundary conditions. The evaporation and transpiration contributions are computed by means of ETo Calculator by FAO which calculates the potential evapo-transpiration using the Penman-Monteith method (1948) in the last revision by Allen et al (1994). Then the actual evaporation is computed by the code and it depends on the moisture state of the soil, becoming maximum after irrigation and decreasing as the surface dries due to a reduction of the permeability according to the hydraulic conductivity function of the soil. $\boldsymbol{ET_o}$ to compute the potential transpiration contribution is calculated using the input meteorological conditions to which the leaves of the plant are subjected (of the room in which the box is stored and radiation from the agricultural lamp in use) while the ETo to compute the potential evaporation from the soil surface is calculated using the meteorological conditions in the enclosed space on the top of the soil surface of the large scale apparatus. The macroscopic root water uptake scheme proposed by Feddes et al (1978) in which the potential transpiration is distributed over the root zone and reduced to the actual root water uptake according to a stress reduction function α(h) has been adopted in Hydrus 2D. The parameters of the Feddes' model have been chosen from the Hydrus internal database based on studies of Wesseling (1991) and Taylor and Ashcroft (1972). The 2D root spatial distribution function $b(x,z)$ of Vrugt et al (2001) implemented in Hydrus2D has been used to define the spatial variation of RWU in the root zone. Due to the fact that the parameters of Vrugt's model are empirical and they do not necessarily reflect real properties of the root system, their values have been chosen ad hoc in the present case study to obtain the best agreement between observed and simulated data. Hydrus is not able to take into account the crop and root growth and the various growth stages of a plant, but due to the fact that the experiment of the large scale apparatus has a reduced duration in time, the root architecture evolution is assumed unchanged during the simulation of the whole experiment. The VGM has been assumed as the hydraulic model of the simulation because it requires the lower number of parameters and the hysteretic effects, for the reduced duration of the experiment, could be neglected. The reduced number of parameters of the hydraulic model assures to focus the calibration on the 6 parameters which define the root spatial distribution function of Vrugt et al (2001) and the 7 parameters which define the water stress reduction function of Feddes et al (1978). Indeed the whole experiment of PART II of the present thesis has the ultimate goal to investigate the evapo-transpiration contribution of plants and the elaborated models used in the numerical analyses to represent this phenomenon. The parameters of the VGM, of the layers improved soil and potting mix, have been obtained from laboratory tests. The hydraulic parameters of the layer of gravel on the bottom of the large scale apparatus have been obtained using pedo-transfer functions based on soil physical properties. The simulation of the large scale experiment starts with phase 1A, the application of the water level at the base of the apparatus and it ends with the flooding of the plant (phase 4). Initial conditions have been set using the pwp information in the observation points and interpolating in the remaining points of the domain. The space discretization of the numerical model is an unstructured mesh of triangular elements which has been refined in close proximity of the observation points and in proximity of the soil surface, where the calculus can have the greater convergence problems.

In the next chapter 3_3, the attention is focused on the sensitivity analysis performed on the parameters of the elaborated numerical model (hydraulic parameters, parameters of the RWU spatial distribution function and of the water stress function) and on the calibration procedure adopted for each of them.

3_3 SENSITIVITY ANALYSIS AND CALIBRATION OF MODEL PARAMETERS

3_3_1 Sensitivity analysis of the model parameters

Before delving into the calibration process, it is relevant to conduct a sensitivity analysis that will reveal the significance of each model parameter on the model output so that the calibration can focus on key parameters. To do so, a one-at-a-time (OAT) sensitivity analysis was performed.

3_3_1_1 Sensitivity analysis of the hydraulic parameters (VGM)

As per the Infrasafe project, the sensitivity coefficient $S(t, b_j)$, defined as a quantitative indicator of change of a model output due to a change of one input parameter of +/-1% (Abbasi, 2015; Simunek et al., 1998), was used for the present analysis. $S(t, b_j)$, is calculated as follows:

$$S(t, b_j) = \frac{|Y(b_j + \Delta b_j e_j) - Y(b_j)|}{Y(b_j)} \quad \text{[equation 77]}$$

where $S(t, b_j)$ is the change in the variable Y corresponding to a 1% change in the parameter b_j; Δb_j=0.01 b_j; e_j is the j_{th} unit vector. For the present model output Y could be water content or pressure head.

The influence of the six hydraulic parameters (VGM) for the improved soil and potting mix, and of the saturated permeability Ks of the gravel was investigated changing their values one-at-a-time of +/-1% their initial value. The output of the model is the value of pressure head at observation point 5 (30 mm on the side of the root bulb). The model was run using the hydraulic parameters in Table 78 (chapter §3_2_4), the parameters of the stress response function of Feddes et al (1978) in Table 76 (chapter §3_2_3), the parameters of the root water uptake spatial distribution function of Vrugt et al (2001) in Table 77 (chapter §3_2_3) and the initial conditions as described in chapter §3_2_5.

In Table 79 the lower (in negative sign) values of the sensitivity coefficient measured as a result of a change of the input hydraulic parameter by +/- 1% are reported. In Figure 68 for each parameter, the evolution of the sensitivity coefficient in time is presented. It can be seen that after a period of 90 days without relevant variations, the coefficient of sensitivity increases almost linearly between 90-210 days and decreases afterwards. This is because it takes about 90 days for the wetting front to reach point 5, after that a slow saturation starts and it is accompanied by a reduction of the influence of the hydraulic parameters. At day 210 the total saturation of the observation point 5 is reached. The θs and α parameters of the improved soil are the only exceptions as they show the opposite trend (increase from 0-90 days then decrease up to day 210).

The same trend was observed for the 1D column experiment, relevant sensitivity coefficients were observed only in the simulation period between the arrival of the hydraulic stimulus and the full saturation of the observation point considered. It is easily observable that the sensitivity of the hydraulic parameters in the 1D filtration experiment is greater than the one observed for the large-scale apparatus experiment (f.e θs has a sensitivity coefficient close to -30%, the n parameter close to -13%). This is symptom of a better predisposition of the 1D filtration model to hydraulic parameters calibration.

It can be seen that the only cases where a relevant variation in the sensitivity coefficient has not been observed is for α and θs parameters of the potting mix (both the sensitivity coefficients between 0 and -1)

This observation means that the calibration programme should apply to all parameters but these two above-mentioned. It is relevant to highlight that the OAT sensitivity analysis has the major limit of investigate the sensitivity of single parameters but not of groups of parameters, this implies that it is legitimate to exclude a parameter with a low sensitivity from calibration performed on the single parameter but not from a calibration performed on a group of parameters.

Table 79: The table reports the greater (negative) values of the sensitivity coefficient measured as a result of a change of the input hydraulic parameter by +/- 1%. (*) the sensitivity coefficient is not calculated for the θr parameter of the potting mix because the initial parameter θr is equal to 0,00 so it is not possible to compute a variation of +/-1%. (**) the sensitivity coefficient is not calculated for the SWRC parameters of the gravel because for the whole duration of the simulation the material is in full saturated conditions. In bold the relevant values of the sensitivity coefficient (below -1%).

Variation of input parameter [%]	l	θ_r	θ_s	α	n	K_s
IMPROVED SOIL +1%	**-1.66**	-0.61	**-2.72**	**-1.52**	**-1.60**	**-5.08**
IMPROVED SOIL -1%	-0.58	**-2.35**	**-2.72**	**-2.72**	**-0.29**	**-1.55**
POTTING MIX +1%	**-1.33**	(*)	-0.34	**-1.73**	-0.95	-0.52
POTTING MIX -1%	**-1.53**	(*)	**-2.75**	**-3.82**	-0.60	-0.23
GRAVEL +1%	(**)					-0.55
GRAVEL -1%						**-3.28**

3_3_1_2 Sensitivity analysis of the water stress function parameters and of the RWU spatial distribution function parameters

A sensitivity analysis of the parameters of the water stress function α(h) and of the RWU spatial distribution function has been carried out in order to identify which parameters influence the most the model outputs.

The parameters tested for the water stress response function α(h) of Feddes et al (1978) are:

- h_1 (value of pressure head below which roots start to extract water from the soil)
- h_2 (value of pressure head below which roots extract water at the maximum rate)
- $h_{3\ high}$ (value of pressure head below which roots can no longer extract water at the maximum rate. A value of T_{p1} has been assumed)
- $h_{3\ low}$ (value of pressure head below which roots can no longer extract water at the maximum rate. A value of T_{p2} has been assumed)
- h_4 (value of pressure head below which root water uptake stops (assumed usually equal to the wilting point)
- T_{p1} (potential transpiration rate, default value 0.5 cm/day)
- T_{p2} (potential transpiration rate, default value 0.1 cm/day)

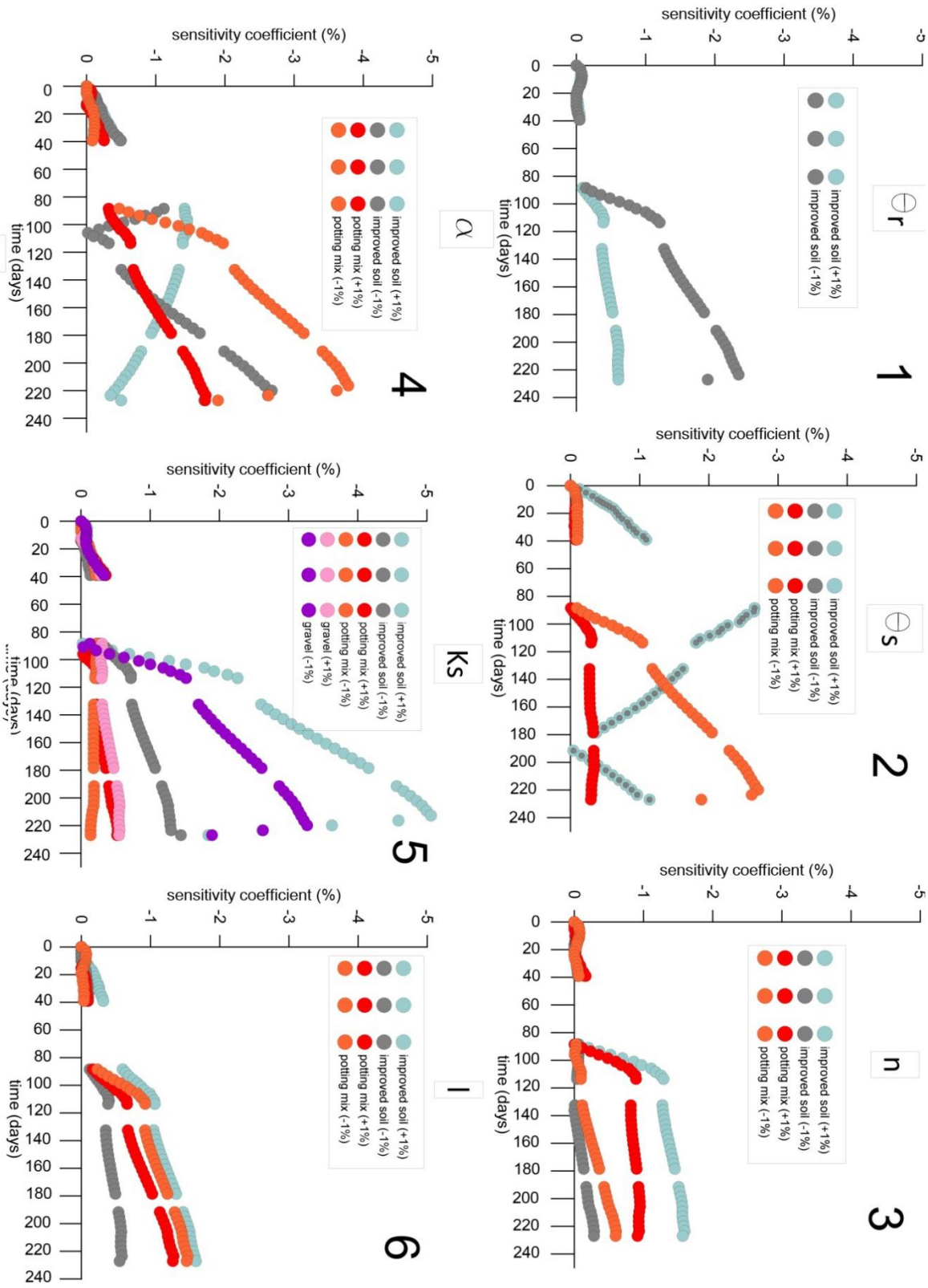


Figure 68: Results of the OAT sensitivity analysis of the hydraulic parameters of the improved soil, potting mix and gravel: evolution of the sensitivity coefficient with time. Time zero on the x-axis corresponds to the beginning of phase 1A (see section 3_6).

An explanatory image of the parameters to which we refer is in Figure 69. For more information on the significance of the Feddes' parameters see chapter §2_1_2_5. The OAT sensitivity analysis has been applied to the model output in the observation points close to the root bulb (from 1 to 11), located in the side of the root bulb (from 1 to 5), below the root bulb (from 6 to 10) and radially from the bulb (11). In Figure 67 the exact location of each observation point is reported. It is interesting at this stage to investigate not only the significance of the different parameters but also how the significance changes in the different parts of the domain. Therefore, the calibration process should not only focus on the identification of the key parameters but also rely on simulation outputs at most sensitive observation points.

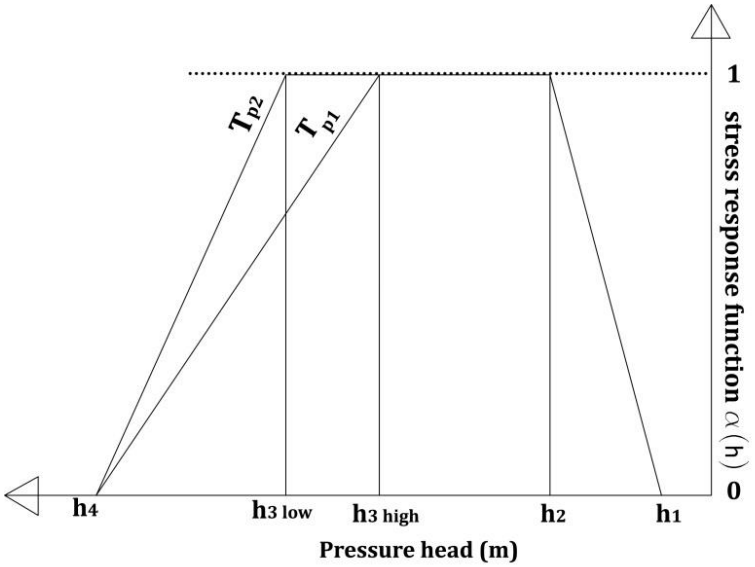


Figure 69: Graphical representation of the stress response function of Feddes et al (1978) with indication of all its parameters.

In Table 80 the minimum (in negative sign) sensitivity coefficient for each parameter of the Feddes et al (1978) model has been reported for each considered observation point. In bold the values of the relevant sensitivity coefficients (values greater than +/-1%). As it is possible to observe only 4 out of 11 observation points show a relevant sensitivity to model parameters change: observation points 5 and 10 (30 mm on the side and below the root bulb, respectively) and 7, 8 below the root bulb (180 and 130 mm from the bulb). The higher influence on the model output is shown by the parameters h_{3high} and h_{3low}, while T_{p1} and T_{p2} have no influence on the simulation output. In Figure 70 the trend in time of the sensitivity coefficients in the observation points 5 and 10 is shown. For observation point 5 the sensitivity coefficient shows small variations up to day 90 when the arrival of the hydraulic stimulus (upward water gradient) is observed, then an increase up to day 220 (full saturation) then a rapid decrease. Observation point 10 shows the opposite trend for h_4 and h_{3low} (increase up to day 90 then a decrease). With great probability this is due to the fact that the wetting front has an earlier influence on observation point 10 with respect to observation point 5 located farther from the water table.

	h_1 (+1%)	h_1 (-1%)	h_2 (+1%)	h_2 (-1%)	h_{3high} (+1%)	h_{3high} (-1%)	h_{3low} (+1%)	h_{3low} (-1%)	h_4 (+1%)	h_4 (-1%)	T_{p1}-T_{p2} (+/-1%)
1	-0.14	-0.14	-0.19	-0.21	-0.21	-0.21	-0.21	-0.14	-0.64	-0.64	0
2	-0.10	-0.10	-0.15	-0.16	-0.17	-0.18	-0.18	-0.11	-0.23	-0.23	0
3	-0.07	-0.06	-0.08	-0.10	-0.11	-0.11	-0.11	-0.11	-0.17	-0.17	0
4	-0.05	-0.02	-0.05	-0.11	-0.08	-0.08	-0.08	-0.05	-0.17	-0.27	0
5	-0.02	-0.04	-0.82	-0.82	**-2.45**	**-2.51**	**-2.51**	**-3.11**	**-2.59**	**-5.14**	0
6	0.00	0.00	-0.66	-0.66	-0.66	-0.65	-0.65	-0.66	-0.66	-0.66	0
7	**-2.78**	**-2.78**	**-2.78**	**-2.78**	**-2.70**	**-2.78**	**-2.78**	**-2.78**	**-2.78**	**-2.78**	0
8	**-15.6**	**-15.26**	**-13.89**	**-15.00**	**-20.45**	**-16.98**	**-16.98**	**-14.80**	**-16.33**	**-29.99**	0
9	-0.02	-0.03	-0.03	-0.03	-0.05	-0.05	-0.05	-0.04	-0.14	-0.14	0
10	-0.82	-0.82	**-1.59**	**-1.59**	**-1.88**	**-1.85**	**-1.85**	**-3.05**	**-1.30**	**-7.07**	0
11	-0.68	-0.66	-0.66	-0.68	-0.65	-0.66	-0.66	-0.67	-0.68	-0.66	0

Table 80: Minimum (in negative sign) sensitivity coefficient obtained from the OAT sensitivity analysis performed on each parameter of the Feddes et al (1978) model and in each observation point close to the root bulb (1 to 11).

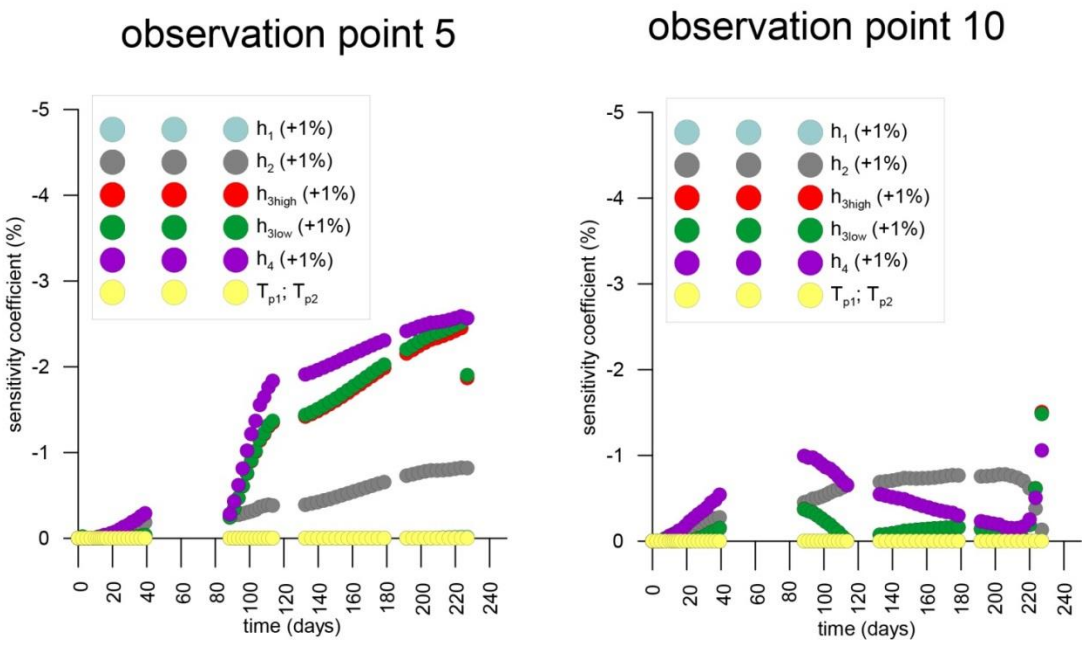


Figure 70: Evolution in time of the sensitivity coefficient calculated for each parameter of the Feddes et al (1978) model in the observation point 5 and 10 (30 mm on the side of the root bulb and 30 mm below the root bulb, respectively). Time zero on the x-axis corresponds to the beginning of phase 1A (see section 3_6).

The parameters tested for the root water uptake spatial distribution function of Vrugt et al (2001) are:

- X_m [L]: Maximum rooting length
- Z_m [L]: Maximum rooting depth
- p_z: Empirical parameter for the vertical direction
- p_r: Empirical parameter for the horizontal direction
- z^*[L]: Depth of maximum intensity
- $x^*[L]$: Radius of maximum intensity

The equation of the root water uptake spatial distribution function of Vrugt et al (2001) is provided again for ease of reading.

$$b(x,z) = (1 - \frac{z}{Z_m})(1 - \frac{x}{X_m})e^{(\frac{p_z}{Z_m}|z^* - z| + \frac{p_r}{X_m}|x^* - x|)} \quad \text{[equation 78]}$$

The model output used is the pressure head at observation points close to the root bulb (see Figure 67 for observation points location). Observation points 1, 2, 3, 4, 5, 11 have been chosen in order to investigate the sensitivity of the parameters moving away from the root bulb on the radial direction and 6, 7, 8, 9 and 10 on the vertical direction. As done for the parameters of the Feddes model, in Table 81 the minimum (in negative sign) sensitivity coefficient is reported for each parameter of the Vrugt et al (2001) model and for each observation point. As could be seen, observation points 5, 10, 7, 8 show the greatest sensitivity as observed also for the Feddes' parameters. Looking closely to the observation points from 1 (230mm from the root bulb, on the side) to 5 (30 mm from the root bulb, on the side), the values of the sensitivity coefficient increase (in negative sign). A more complex trend is shown by the observation points below the root bulb with a reduction of the coefficient passing from sensor 10 (30 mm from the bulb) to sensor 9 (80 mm from the bulb) but then a peak of sensitivity is observed in correspondence of observation point 8.

The parameters that have the most influence on the outputs are the maximum rooting lengthX_m, the depth of maximum intensity Z_m and with a lower relevance the depth of maximum intensity z^* and the radius of maximum intensity x^*. In contrast, the output seems to be little sensitive to the empirical parameter for vertical and horizontal direction, p_z and p_r.

In Figure 71 the trend of the sensitivity coefficients in time is presented for observation point 5 (30 mm on the side of the root bulb) and 10 (30 mm below the bulb). The coefficient shows in time a slow increase up to day 220 (flooding of the apparatus).

	p_r $+1\%$	p_r -1%	p_z $+1\%$	p_z $+1\%$	X_m +1%	X_m -1%	Z_m +1%	Z_m -1%	z^* +1%	z^* -1%	x^* +1%	x^* -1%
1	-0.82	-0.90	-0.16	-0.13	-0.23	-0.21	-0.18	-0.13	-0.16	-0.19	-0.21	-1.25
2	**-1.32**	**-1.40**	-0.19	-0.13	-0.18	-0.16	-0.17	-0.13	-0.24	-0.31	-0.16	**-2.19**
3	**-2.62**	**-2.79**	-0.41	-0.34	-0.18	-0.28	-0.40	-0.31	-0.54	-0.75	-0.28	**-5.53**
4	**-4.25**	**-4.49**	-0.85	-0.83	-1.91	-2.10	-0.97	-0.73	**-1.09**	**-1.70**	**-2.10**	**-13.55**
5	**-6.37**	**-4.43**	**-1.17**	**-1.17**	**-7.08**	**-6.25**	**-2.95**	**-2.22**	**-1.04**	**-4.17**	**-6.25**	**-23.08**
6	-0.34	-0.34	-0.34	-0.28	-0.34	-0.28	-0.66	-0.66	-0.34	-0.23	-0.28	-0.28
7	**-2.70**	**-2.70**	**-2.78**	**-2.70**	**-2.94**	**-3.03**	**-3.13**	**-3.03**	**-2.70**	**-3.03**	**-3.03**	**-3.03**
8	**-7.19**	**-10.18**	**-13.26**	**-8.13**	**-2.37**	**-48.25**	**-107.96**	**-83.04**	**-36.01**	**-74.08**	**-48.25**	**-25.09**
9	-0.68	-0.69	-0.62	-0.62	**-2.98**	**-3.16**	**-16.27**	**-11.07**	**-1.97**	**-1.92**	**-3.16**	**-1.66**
10	**-3.58**	**-5.42**	**-3.99**	**-6.81**	**-10.53**	**-11.21**	**-8.12**	**-8.57**	**-3.86**	**-8.67**	**-11.21**	**-8.75**
11	-0.68	-0.69	-0.66	-0.58	-0.68	-0.69	**-1.19**	-0.91	-0.68	-0.68	-0.69	-0.69

Table 81: Minimum (in negative sign) sensitivity coefficient obtained from the OAT sensitivity analysis performed on each parameter of the Vrugt et al (2001) model and in each observation point close to the root bulb (observation points 1 to 11).

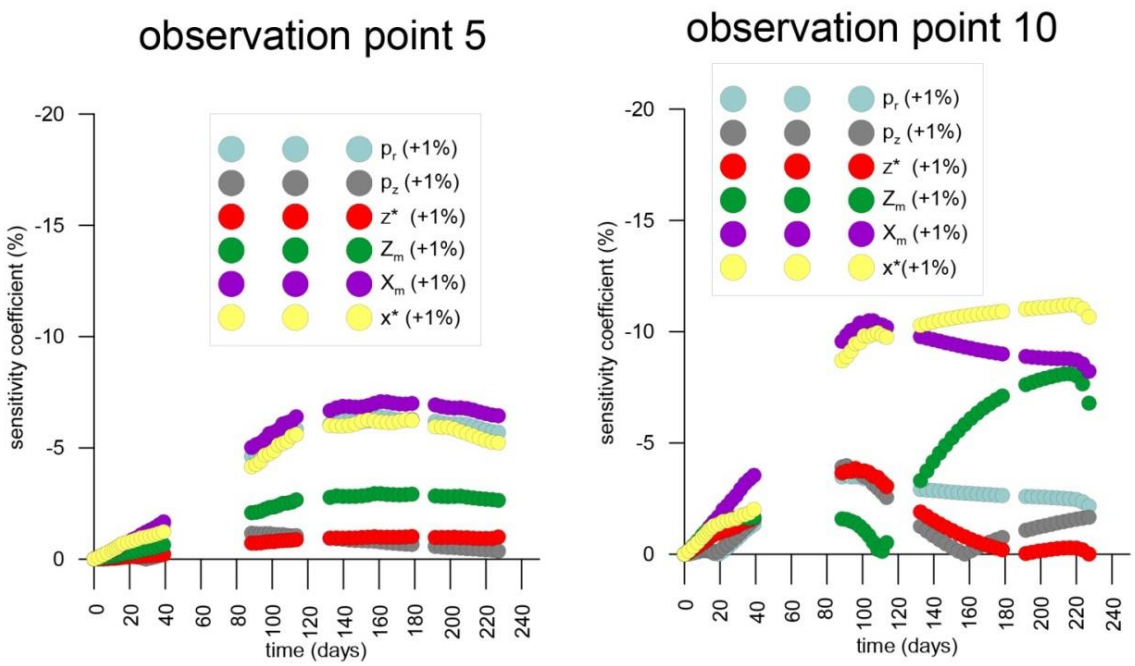


Figure 71: Evolution in time of the sensitivity coefficient calculated for each parameter of the Vrugt et al (2001) model in the observation point 5 and 10 (30 mm on the side of the root bulb and 30 mm below the root bulb, respectively). Time zero on the x-axis corresponds to the beginning of phase 1A (see section 3_6).

3_3_2 Calibration of model parameters

3_3_2_1 Calibration of potential RWU pattern from pore pressure readings

As discussed in section 2_1_2_2, the RWU model used in this thesis is based on the model by Raats (1974) which defines the spatial root distribution in the vertical direction and it is extended with the radial component by Vrugt et al (2001). The spatial root distribution function of Vrugt et al (2001) implemented in Hydrus 2D requires the definition of 6 parameters: $Z_m, X_m, p_z, p_r, z^*, x^*$(see equation 17 in chapter §2_1_2_2). Hydrus 2D has implemented inverse analysis only for the optimization of the hydraulic parameters (SWRC and hydraulic conductivity function) and not for the optimization of the Vrugt's parameters of the spatial distribution function. Thus, only a trial-and-error procedure of calibration could be used for Vrugt's parameters. In order to ease the calibration of the RWU pattern i.e. the definition of the local dimensionless root distribution parameter $b(x,z)$ in each point of the domain affected by the root uptake, another calibration approach is used. The local value of $b(x,z)$ leads the potential root uptake in that node as indicated in equation 12 (chapter §2_1_2_1). Equation 12 is reported here again for a sake of clarity:

$$S_p(x,z,t) = b(x,z,t)L_t T_p(t) \qquad \text{[equation 79]}$$

Where L_t [L] is the width of the soil surface associated with the plant transpiration [L] and T_p [LT^{-1}] the potential transpiration.

The adopted calibration procedure focuses on the definition of the local dimensionless root distribution parameter $b(x,z)$ in each of the observed nodes (where sensors have been installed) using the spatial distribution of pore pressure given by the Teros 21 suction sensors installed around the root bulb.

The local value of $b(x,z)$ is adjusted by the user in order to obtain the best possible fit between the simulated and observed decay of suction with time. Hydrus allows to set the value of $b(x,z)$ graphically, node by node, as an alternative to use the spatial distribution of Vrugt et al (2001).

The modeller sets, using a trial and error procedure, the local dimensionless root distribution parameter $b(x,z)$ in the observation points around the bulb trying to minimize the differences between observed decay of water potential in time and simulated one. To extrapolate the unknown values of $b(x,z)$ in the other nodes of the domain an interpolation using the known values in the observed nodes is performed and assuming the following hypothesis:

- the maximum value of $b(x,z)$ is located in exact centre of the root bulb and decreases rapidly away from the bulb centre, consistent with physical observations reported by other researchers (Knight, 1999; Doussan et al, 2006)
- A symmetrical (with respect to the vertical axis) RWU pattern is assumed.

The observation points considered are 1, 2, 3, 4, 5 and 11 on the side of the plant; 8, 9 and 10 below the plant. These points were selected because the effect of the root bulb is very clear at these points. The simulations were conducted over a period of 6 days (first 6 days of Phase 1A of the experiment, see section 3_6 for the experimental program), for the following reasons:

- time zero of the simulation corresponds to right after the last irrigation of the plant from the top in phase 0 (prior to the application of the water level at the lower boundary of the box). At that time, the root bulb is close to saturation and the plant does not experience water stress, as observed in the sensitivity analysis of the parameters governing the water stress function of the Feddes' model in Figure 70 in the first six days of the simulation (sensitivity coefficient close to zero). For this reason, in the first part of the simulation, the potential uptake coincides with the actual uptake (no reduction of the uptake due to environmental stress). Keep separated the influence of the stress response function α(h) from the effects on the potential transpiration of the RWU pattern is hence ideal for the calibration of the RWU.
- As the sensitivity analysis of the hydraulic parameters shows in Figure 68, the hydraulic parameters of the VGM have a very limited influence on the model during the first period of the simulation, which is governed almost completely by the spatial root uptake parameters and is hence ideal for the calibration of the RWU.
- Over such a short period, the effect of redistribution of the water profile from the irrigation point can be neglected (Coelho and Or, 1996).

In Figure 72, the calibrated RWU pattern, expressed in terms of local dimensionless root distribution parameter $b(x,z)$ and using in situ observation points, is presented. Numbers in black at the observation points are the experimental $b(x,z)$ values from which the spatial distribution of the RWU has been reconstructed.

In Figure 74, the experimental and simulated evolution of suction in time for the different observation points over the period of 6 days is compared, for the initial reconstruction of the RWU spatial distribution. As could be observed, a satisfactory match is obtained for almost all the observation nodes.

A similar approach was employed using Vrugt's model to establish the initial RWU pattern (see Table 82 with the chosen parameters of the Vrugt's model). The parameters of the Vrugt's model have been tuned in order to reduce as much as possible the discrepancy between observed values and simulated values. The

RWU pattern obtained after the ''trial and error'' calibration procedure using the Vrugt's model can be observed in Figure 73.

The evolution of suction with time is presented in Figure 75. It shows the comparison between the observation points (blue dots), the pwp trend using the Vrugt's model (red dots) and the pwp using the RWU pattern elaborated without adopting any spatial distribution model but only the experimental information in the observation points (green dots). Overall, the outcome is not satisfactory because Vrugt's function imposes many constraints on the RWU pattern and for this reason it is not possible to tune adequately the outcome of the model to the observation points in the observation nodes. Vrugt's function is not able to reconstruct the RWU spatial distribution with high accuracy.

To confirm the fact that the RWU pattern elaborated without adopting any spatial function is better than the one elaborated using the Vrugt's model, the goodness of fit between experimental observations and simulations was quantified using a set of efficiency metrics ($IA_{j=1}$, IA_j, IA_{rel}, NSE, NSE_j, NSE_{rel}) and accuracy metrics (R^2, MAE, RMSE) (Table 83).

The metrics consistently show that the RWU spatial distribution elaborated using only the experimental data and any spatial distribution function better represents the experimental observations.

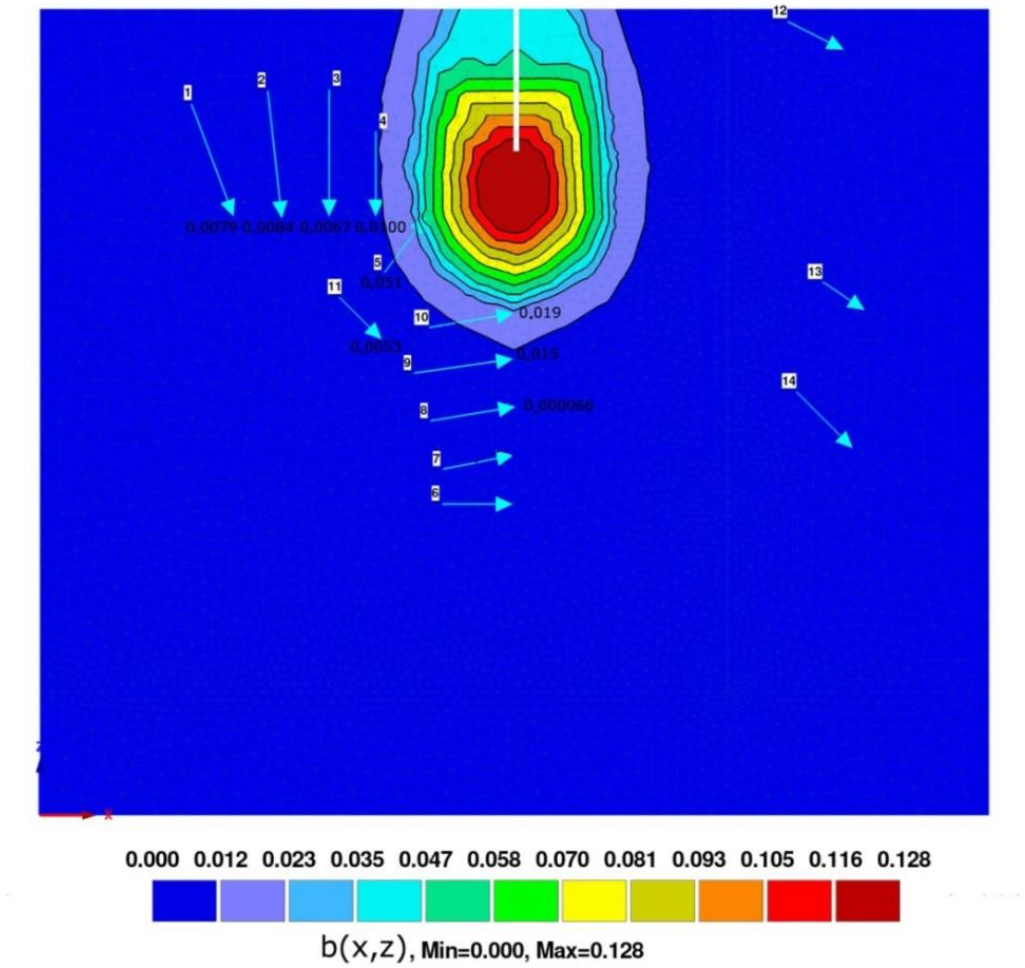


Figure 72: RWU spatial distribution obtained tuning the b(x,z) parameter in each of the observation nodes in order to obtain the best fit between simulated and observed values, in the first 6 days after the starting time of the simulation (phase 1A, after the last irrigation along the plant stem).

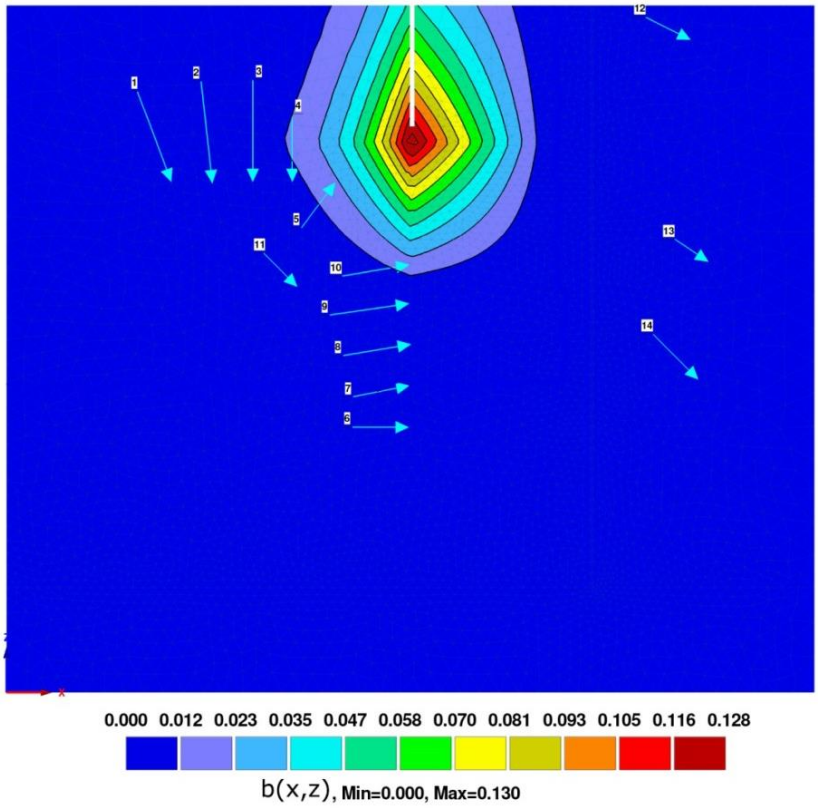


Figure 73: The RWU spatial distribution function of Vrugt et al (2001). Parameters are chosen in order to obtain the best match with the observed data in the observation nodes (where sensors have been installed).

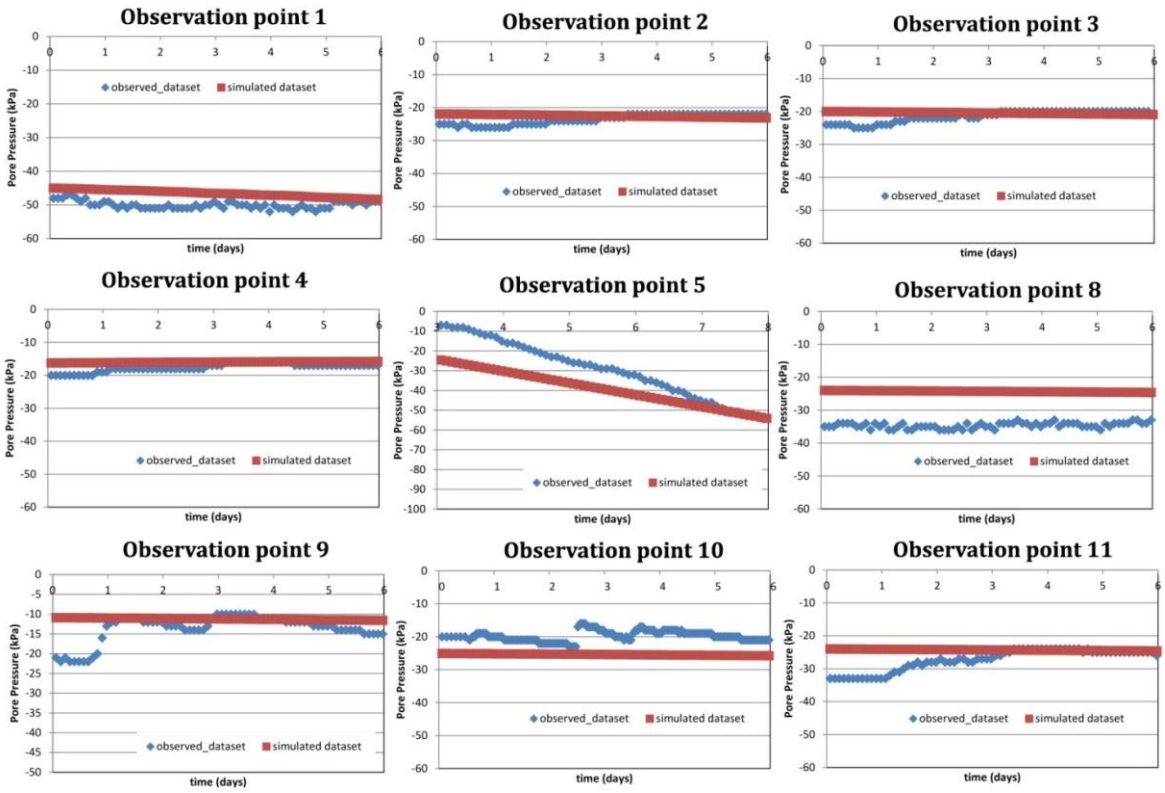


Figure 74: The pore pressure trend in time (first 6 days) of the observation points (blue points) and of the simulated points (red points) which helped defining the b(x,z) values in the observation nodes. These values of b(x,z) helped reconstructing the RWU spatial distribution of the investigated plant without the use of a spatial function.

Table 82: The parameters of the spatial distribution function of Vrugt et al (2001) used to elaborate the RWU pattern in Figure 73

Parameters of Vrugt (2001)'s model for vertical distribution		
Maximum rooting depth Z_m	Depth of maximum intensity z^*	Empirical parameter p_z
0.36	0.17	3.5
Parameters of Vrugt (2001)'s model for horizontal (radial) distribution		
Maximum rooting radius X_m	Radius of maximum intensity x^*	Empirical parameter p_r
0.675	0.5	13.4

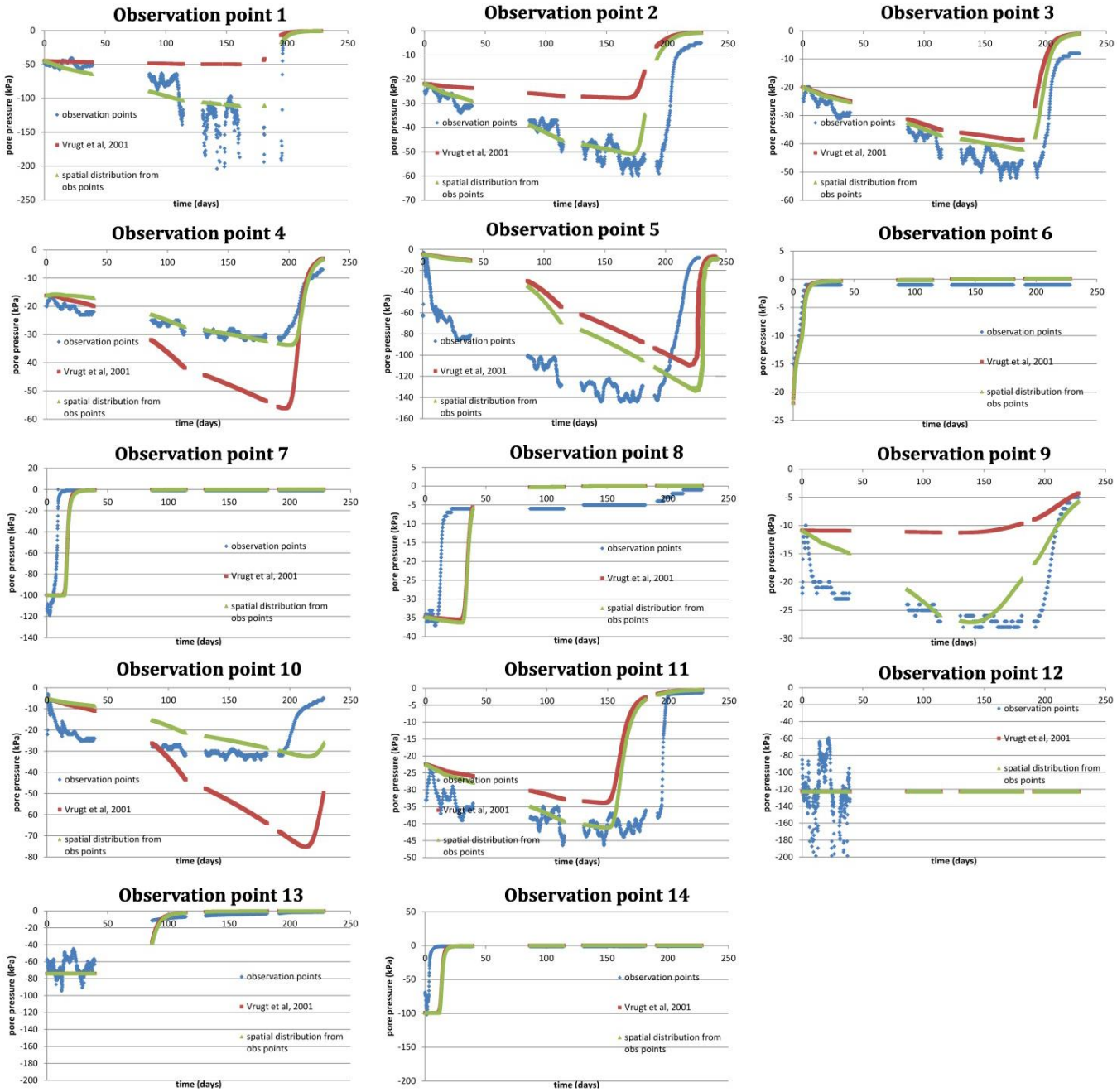


Figure 75: The pore pressure trend in time (full simulation period) using the Vrugt's RWU spatial distribution (red dots), the spatial distribution using the observed dataset (green dots) and the observation points (blue dots).

Table 83: Quantitative comparison between simulated – observed values using the RWU spatial distribution function of Vrugt's et al (2001) and using the spatial distribution elaborated starting from experimental data in the observation points (first 6 days of the simulation) and no use of spatial functions.

best fit	1	1	1	1	1	1	1	1	1	0	0
	Observation point 1										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.572	0.657	0.572	0.901	0.262	0.111	0.744	0.631	0.484	1.459	9.905
RWU spatial distribution from experimental data	0.801	0.929	0.801	0.934	0.612	0.762	0.781	0.806	0.764	0.766	5.127
	Observation point 2										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.476	0.632	0.476	0.703	-0.100	-0.510	-0.220	0.450	0.338	12.945	17.250
RWU spatial distribution from experimental data	0.787	0.876	0.787	0.839	0.534	0.446	0.281	0.767	0.633	5.477	10.449
	Observation point 3										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.666	0.862	0.666	0.795	0.346	0.435	0.160	0.673	0.760	7.178	9.525
RWU spatial distribution from experimental data	0.747	0.924	0.747	0.841	0.493	0.683	0.342	0.746	0.878	5.562	7.134
	Observation point 4										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.328	0.373	0.328	0.667	-1.023	-2.037	-0.612	-0.012	0.001	16.389	22.046
RWU spatial distribution from experimental data	0.499	0.438	0.499	0.815	-0.049	-0.996	0.342	0.475	0.007	8.497	17.871
	Observation point 5										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.362	0.574	0.362	-4.020	-0.318	-0.562	-17.419	0.341	0.160	52.609	59.319
RWU spatial distribution from experimental data	0.419	0.621	0.419	-5.975	-0.206	-0.414	-25.019	0.397	0.170	48.128	56.429
	Observation point 6										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.288	0.476	0.288	0.570	-1.473	-3.116	-2.379	-0.237	0.719	12.216	13.105
RWU spatial distribution from experimental data	0.588	0.791	0.588	0.761	-0.014	-0.035	-0.182	0.493	0.627	5.010	6.573
	Observation point 10										
	$IA_{j=1}$	IA	IAj	IArel	NSEj	NSE	NSErel	IA'	R^2	MAE	RMSE
Vrugt's RWU spatial distribution	0.130	0.226	0.130	-3.232	-0.888	-1.672	-13.607	0.056	0.012	10.675	12.264
RWU spatial distribution from experimental data	0.193	0.238	0.193	-3.828	-2.849	-13.437	-90.471	-0.925	0.050	21.771	28.506

3_3_2_2 Calibration of the compensatory RWU mechanism

The compensatory RWU mechanism takes into consideration the fact that the plant may respond non uniformly over the root zone to water stress. The reduced amount of water uptake in the most water stressed part of the root zone is compensated by an increase of uptake from the less stressed parts (Simunek and Hopmans, 2009; Jarvis, 1989).

The degree of compensation is governed by the critical water stress index (introduced in section 2_1_2_3) which takes into consideration that the RWU at a certain depth in the root system is influenced by pwp and uptake at other depths.

There exists no clear way to experimentally measure the critical water stress index value so researchers have to resort to fitting numerical responses to experimental ones in order to estimate it.

Here, the critical water stress index was incrementally increased from 0.6 to 1 using the previously calibrated RWU with the objective to identify which value would result in the best match with the experimental evolution of suction in time at the different observation points. A water stress index equal to 1 means no compensatory mechanism considered.

Figure 76 shows the results of simulations with varying critical stress index as well as experimental values. Considering the compensatory mechanism means that the plant will extract more water from the wetter part of the root domain to compensate the reduction in RWU in the more water stressed parts, resulting in higher suction values, which can clearly be seen in the observation points of Figure 76.

Table 84 reports the maximum values of suction (minimum negative pwp value) for a water stress index of 0.6 and 1 as a function of distance to the stem (second column). The third column of Table 84 reports the difference between the minimum pwp value in the configuration with ω equal to 1 (no compensatory mechanism) and in the configuration with ω equal to 0.6. For what concerns the observation points on the side of the bulb; observation points show almost the same increments in suction between the two configurations due to a lower b(x,z) value but an higher availability of water for the more external observation points and an higher b(x,z) value but a lower availability of water for the points closer to the bulb. For what concerns the observation points below the bulb, observation point 10 shows the higher increment in suction (104 kPa) passing from a critical water stress index of 1 to 0.6, due to a very high $b(x,z)$ value. Observation point 5 (30 mm on the side of the bulb) shows a lower increment compared to observation point 10 due to the fact that -160 m has been set as wilting point and the roots can no longer extract water below this point.

The goodness of fit between observed and simulated data has been assessed by a series of efficiency metrics ($IA_{j=1}$, IA_j, IA_{rel}, NSE, NSE_j, NSE_{rel}) and accuracy metrics (R^2, MAE, RMSE), see Table 85.

The critical water stress values returning the best goodness of fit (observed-simulated pwp) is indicated in red in the first column. In most cases, the best fit is obtained for critical water stress values of 1 that means no compensatory ability of the considered plant. As reported by Cai et al (2017) and Simunek and Hopmans (2009), the critical stress index is relatively high for cultural plants compared to natural plants because cultural plants have a low (or zero) ability to compensate for water stress.

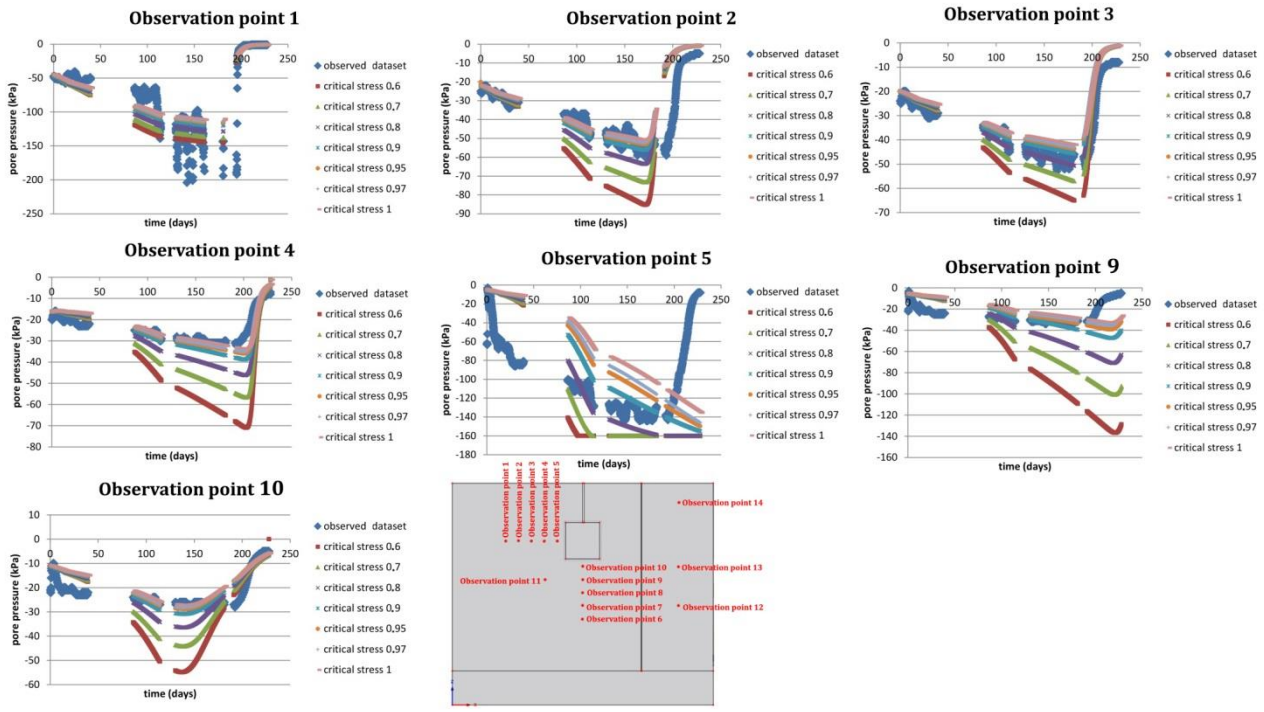


Figure 76: Comparison between the performance of the simulation in the observation points changing the critical water stress index ω_c (from 0.6 to 1).

	Distance from the root bulb (mm)	Min pwp (ω_c=0.6)	Min pwp (ω_c=1) no compensatory mechanism	Difference pwp ω_c=0.6 and ω_c=1
Observation point 1	230-horizontal direction	-145.9	-111.74	**-34.18**
Observation point 2	180-horizontal direction	-85.21	-50.68	**-34.53**
Observation point 3	130-horizontal direction	-65.03	-42.03	**-23.00**
Observation point 4	80-horizontal direction	-70.95	-33.69	**-37.26**
Observation point 5	30 -horizontal direction	-160.04	-135.45	**-24.59**
Observation point 9	80-vertical direction	-54.91	-27.14	**-27.76**
Observation point 10	30-vertical direction	-136.76	-32.52	**-104.24**
Observation point 11	80-vertical direction 80-horizontal direction	-29.98	-27.25	**-2.73**

Table 84: The minimum negative pwp value is reported for each observation point in the two cases: adoption of the critical water stress index ω=1 and ω=0.6. In the last column the difference between the two min pwp values.

best fit	1	1	1	1	1	1	1	1	1	0	0
	Observation point 1										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0.801	0.929	0.801	0.934	0,612	0,762	0.781	0.806	0.764	0,766	5.126
crit stress 0.97	0.800	0,930	0.800	0.931	0.607	0.761	0.766	0.803	0.765	0,778	5,141
crit stress 0.95	0.799	0.930	0.799	0.929	0.602	0.758	0.755	0.801	0.766	0,787	5.165
crit stress 0.90	0.796	0.930	0.796	0.923	0.586	0.747	0.722	0,793	0.767	0,817	5,280
crit stress 0.80	0.779	0.923	0.779	0.905	0.528	0.696	0.626	0.764	0,765	0.934	5.796
crit stress 0.70	0.752	0.909	0.752	0.883	0.443	0.609	0.496	0.722	0.762	1.100	6.569
crit stress 0.60	0.727	0.893	0.727	0.859	0.358	0.498	0.340	0.6792	0.7573	1.2683	7.4380
	Observation point 2										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0,787	0,876	0.787	0.839	0.534	0.446	0.281	0.767	0.633	5,477	10.449
crit stress 0.97	0.791	0.881	0.791	0,844	0.536	0.452	0.284	0,768	0.639	5.460	10,386
crit stress 0.95	0.792	0.883	0.792	0,847	0.533	0.453	0.283	0.766	0.643	5.496	10.377
crit stress 0.90	0.787	0,887	0.787	0.853	0.507	0.440	0.274	0.754	0.653	5.795	10.508
crit stress 0.80	0.740	0.878	0.740	0.860	0.348	0,296	0,193	0.674	0.671	7,671	11.778
crit stress 0.70	0,673	0.842	0.673	0.854	0.101	-0.109	-0.025	0.551	0.685	10.576	14.778
crit stress 0.60	0.604	0,782	0.604	0.836	-0,215	-0.941	-0,462	0.393	0.695	14.290	19.556
	Observation point 3										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0.747	0.924	0,747	0.841	0.493	0.683	0.342	0,746	0.878	5.562	7.134
crit stress 0.97	0.769	0,933	0.769	0,849	0.531	0.716	0.365	0,765	0.884	5.150	6,755
crit stress 0.95	0,783	0,938	0.783	0.855	0,555	0.737	0.380	0.778	0.888	4.880	6.503
crit stress 0.90	0.817	0.951	0.817	0.867	0,614	0.782	0.413	0,807	0,897	4.230	5.915
crit stress 0.80	0.842	0,965	0.842	0.889	0.643	0.827	0.459	0,822	0.915	3.914	5.264
crit stress 0.70	0.791	0.953	0,791	0,899	0.486	0.736	0.430	0.743	0,929	5,641	6.511
crit stress 0.60	0,723	0.912	0.723	0.893	0.243	0.394	0,266	0.622	0.936	8.299	9.866
	Observation point 4										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0.499	0.438	0.499	0.815	-0.049	-0.996	0.342	0.475	0.007	8.497	17.871
crit stress 0.97	0.496	0.437	0.496	0,812	-0,068	-0.999	0.332	0.466	0.006	8.653	17.885
crit stress 0.95	0.488	0.435	0.488	0.808	-0.094	-1.006	0.319	0.453	0.006	8.864	17.917
crit stress 0.90	0.455	0.429	0.455	0,794	-0.202	-1.047	0.260	0,399	0.006	9,737	18.098
crit stress 0.80	0,373	0.401	0.373	0,723	-0.558	-1,309	-0.069	0.221	0.004	12.623	19,222
crit stress 0.70	0,304	0.344	0.304	0.586	-1.062	-2,053	-0.930	-0.031	0.004	16,704	22.103
crit stress 0.60	0.241	0.268	0.241	0.407	-1,746	-3.698	-2.803	-0.373	0.004	22.244	27,421
	Observation point 5										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0.419	0.621	0.419	-5.975	-0,206	-0.414	-25.019	0.397	0.170	48.128	56.429
crit stress 0.97	0.470	0,653	0.470	-6.797	-0,133	-0,359	-29.527	0.433	0.185	45.220	55,330
crit stress 0.95	0.509	0,679	0,509	-6.983	-0,068	-0,297	-31.244	0,466	0.208	42.613	54,053
crit stress 0.90	0.591	0.723	0.591	-7,012	0,075	-0.209	-33,955	0.537	0.262	36.927	52.181
crit stress 0.80	0.647	0,766	0.647	-6.141	0,101	-0.238	-36.829	0,550	0.373	35.876	52.806
crit stress 0.70	0,621	0,769	0.621	-5,499	-0,042	-0,347	-36.881	0.479	0.427	41.577	55,076
crit stress 0.60	0.602	0.760	0.602	-5,265	-0,129	-0.452	-36.949	0.436	0.435	45.048	57.188
	Observation point 9										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0.588	0.791	0,588	0.761	-0.014	-0.035	-0.182	0,493	0.627	5.010	6.573
crit stress 0.97	0.608	0.803	0.608	0.771	0.015	0.005	-0,154	0.507	0.622	4.865	6.445
crit stress 0.95	0.615	0.810	0.615	0.777	0.019	0.025	-0.141	0.509	0.618	4.848	6.380
crit stress 0.90	0.619	0,821	0.619	0.790	-0.018	0.037	-0.133	0.491	0.608	5.028	6.339
crit stress 0.80	0.582	0,809	0.582	0,795	-0.260	-0,226	-0.316	0.370	0.584	6,221	7,151
crit stress 0.70	0.511	0.734	0.511	0,758	-0.706	-1,272	-1.060	0.147	0.553	8.428	9.738
crit stress 0.60	0.428	0.613	0.428	0.690	-1.361	-3.823	-2.859	-0,181	0.517	11.663	14.187
	Observation point 10										
	$IA_{j=1}$	IA	IA_j	IA_{rel}	NSE_j	NSE	NSE_{rel}	IA'	R^2	MAE	RMSE
crit stress 1	0,131	0.227	0.131	-3.263	-0,865	-1.645	-13.589	0.067	0,010	10,549	12.202
crit stress 0.97	0.178	0.247	0.178	-3.759	-0,819	-1,798	-16.688	0.091	0.018	10,286	12.550
crit stress 0.95	0,241	0.270	0.241	-4.289	-0.796	-2.165	-21.917	0.102	0,031	10,156	13.346
crit stress 0.90	0.277	0,283	0.277	-4,906	-0,996	-3.441	-35.573	0.002	0.057	11,287	15.810
crit stress 0.80	0,211	0.232	0.211	-5.157	-2.301	-10,920	-94.559	-0.651	0.099	18.672	25.901
crit stress 0.70	0.147	0.171	0.147	-4.927	-4.370	-28.683	-211,244	-1,685	0.116	30.368	40,874
crit stress 0.60	0,103	0,126	0.103	-4.600	-7.113	-64.430	-418.439	-3,056	0.116	45.883	60.685

Table 85: Values of metrics used to evaluate the performance of the simulation in the observation points changing the critical stress index ω_c (from 0.6 to 1).

3_3_2_3 Calibration of hydraulic parameters (SWRC and hydraulic conductivity function) through inverse analysis

A calibration was also performed to optimize the hydraulic parameters (SWRC and hydraulic conductivity function parameters) of the two materials (improved soil and potting mix). The hydraulic parameters of the gravel (layer on the bottom of the large-scale apparatus) have not been optimized because the material remains saturated during the whole experiment and the only relevant hydraulic parameter is the saturated permeability. However, Ks is so high (compared to the saturated permeability of the rest of the materials), that could be assimilated to a water reservoir.

The procedure in two phases is the same conducted in section 3_1_2_7 (PART II) and it is simpler than the one applied to the calibration of the numerical model of the embankment section along Secchia River (PART I of the present thesis). The addressed problem is simpler (optimization of the hydraulic parameters of two materials instead of 4, short simulation period, small spatial domain, lower number of performed indirect simulations).

In particular the θr, θs, α, n parameters that define the SWRC are optimized together with the pore connectivity factor l and the saturated permeability Ks which allow from the SWRC function to predict the hydraulic conductivity function HCF (Mualem, 1976). The analytical formulation of the VG's function and Mualem's function defining the SWRC and HCF respectively, is expressed in equations from 72 to 76.

The experimental data chosen for the calibration are those of phases 1A to phase 3 (see section 3_6). All suction data were used for the calibration, both in the control zone and in the zone of the large-scale apparatus subjected to plant evapo-transpiration effects.

A weight of 0.5 has been set to sensors 6, 7, 8, 13 and 14 (as opposed to 1 for all other sensors) because, in the course of the experiment, they end up being below the water level, returning non-meaningful data (the minimum suction value these sensors can read is 9kPa), as such, their contribution to parameters optimization is of less compared to others sensors.

23 simulations were performed to optimize the parameters of the improved soil, 23 to optimize the parameters of the potting mix and 22 for the combined optimization of the parameters of the improved soil and potting mix. 11 simulations were successful (reached convergence) for the improved soil, 23 simulations were successful for the potting mix and 4 simulations were successful for the potting mix combined to the improved soil.

As done previously, different combinations of parameters were optimized: single parameter, pairs of parameters, triplet of parameters, quatrain of parameters to all the parameters of the VG model together. The calibration protocol in two phases presented and applied in chapter §3_1_2_7 to the 1D column model experiment, has also been applied to the large-scale apparatus experiment.

A calibration programme in two phases has been applied to the output of the indirect simulations in order to evaluate quantitatively their performance by means of metrics/indices/coefficients/properties. As already reported in chapters §2_2 and §3_1_2_7 (PART II), the first phase of the calibration focuses on the simulation output which is compared to the observation data and a set of metrics/indices are used to evaluate quantitatively the fit between observed-simulated datasets. The second phase of the calibration

focuses on the last iteration of the inverse analysis, on the quality of the prediction of the set of optimized parameters and the accuracy with which the problem equations have been solved.

In the first phase of the calibration, the goodness of fit between observed and simulated data was computed over the whole simulation period using efficiency metrics (the Nash-Sutcliffe Efficiency NSE, the modified form of the Nash-Sutcliffe Efficiency NSE_j, the index of agreement IA, the modified form of the index of agreement IA' and IA_j, the Kling-Gupta Efficiency criteria KGE) and accuracy metrics (the root mean square error RMSE, the mean absolute error MAE, the mean square error MSE, the coefficient of determination R^2, the coefficient of residual mass CRM). Each metric/index has a different weight as in Table 70 (chapter §3_1_2_7). Points are computed for each metric/index as in equations 52 and 53. For more information on the adopted procedure see chapter §3_1_2_7.

The properties/coefficients of the second calibration phase are: the number of iterations performed in the indirect problem (NoI), the mass balance error (MBE), the confidence interval of an optimized parameter (CI), the standard error coefficient of the optimized parameters (SEC), the correlation coefficient of the combination of parameters (CC) and the sum of squares residuals (SSQ). Each coefficient/property has a different weight as in chapter §3_1_2_7. Points are computed for each coefficient/property as in equations 52 and 53. For more information on the adopted procedure see chapter §3_1_2_7.

In Table 86 the scores attributed to the first and second phases of the calibration program are reported for each indirect simulation; in the last column the final score (sum of the scores of the first and second calibration phases). The indirect simulation that shows the higher score among the performed ones is highlighted in red. The indirect simulation that shows the higher score optimizes the hydraulic parameters Ks and θs of the potting mix while the remaining parameters do not show any changes with respect to the initial values. The higher score of this simulation assures the greatest enhancement of the model output (among the performed indirect simulations) in the considered observation points which is the final purpose of the calibration program. In Figure 77 the evolution of suction with time in the considered observation points is shown. The simulated dataset that uses the set of initial parameters (green dots) is compared to the one that uses the set of optimized parameters (red dots) and the observed dataset (blue dots). As it is possible to observe an enhancement in the model performance is observed only for observation points 5 and 10 (the closer to the root bulb and the potting mix). The same conclusions can be drawn looking at the values of the metrics in Table 87. There are only small differences in goodness of fit for sensors 5 and 10, and close to no difference for the rest of the sensors. In this case study, the main limit of the inverse analysis application when more than one set of hydraulic parameters has to be optimized together, can be observed. The enhancements of the model performance are often localized in one single material and rarely groups of materials optimized together give the greatest enhancement among the performed indirect simulations. This is due to the fact that problems of uncertainty, non-identifiability and correlation of the optimized parameters increase increasing the number of parameters to optimize. The same conclusions were found also in PART I of the present thesis in the calibration results of the bank section along Secchia river.

Table 88 provides the initial set of hydraulic parameters and the final set of calibrated hydraulic parameters. It can be seen that the saturated permeability Ks of the potting mix passes from $8.42*10^{-6}$ m/sec to $7.86*10^{-6}$ m/sec (parameter variation of -6.65%, high confidence limit of 0.57 meaning uncertainties in the determination of the parameter) and θs passes from 0.6614 to 1.757 (parameter

variation of +62%, low confidence limit of 0.0935) and a correlation coefficient of 0.0163 (a very low value of correlation meaning a positive independence in the determination of the parameters).

The low enhancement obtained at the end of the calibration process could be due to a low influence of the SWRC parameters on the trend of suction with time in the considered observation points. In other words, the phenomenon described by the numerical model is more influenced by the RWU spatial distribution parameters and, to a lower extent, by the Feddes' stress function parameters than by the SWRC parameters. The higher sensitivity of the RWU spatial parameters with respect to the other model parameters was observed also in chapter §3_3_1_2.

The calibration program applied to the results of the 1D infiltration experiment has obtained in comparison a greater enhancement in model performance, due to a greater sensitivity of the model to hydraulic parameters changes as shown by the sensitivity coefficients (see Table 67 of chapter §3_1_2_6).

n° simulation	optimized parameters	SCORE 1st PHASE	SCORE 2nd PHASE	SCORE 1st+2nd PHASE
potting mix				
1	α	4.910	2.259	7.169
2	α+n	4.911	2.207	7.117
3	α+θs	4.906	2.305	7.211
4	Ks	4.911	2.282	7.193
5	Ks+α	4.912	2.328	7.239
6	Ks+α+θs	4.910	1.991	6.901
7	Ks+l	4.911	1.974	6.885
8	Ks+l+θs	4.913	2.095	7.008
9	Ks+n	4.911	3.335	8.246
10	Ks+n+alfa	4.910	1.845	6.756
11	Ks+n+θs	4.910	2.023	6.934
12	Ks+θr	4.895	2.780	7.675
13	**Ks+θs**	**4.913**	**2.999**	**7.911**
14	l	4.911	2.475	7.386
15	n	4.911	2.749	7.660
16	n+l+θs	4.912	2.095	7.006
17	n+θs	4.912	2.066	6.978
18	θr	4.911	2.424	7.334
19	θr+α+θs	4.909	2.087	6.996
20	θr+θs	4.911	2.140	7.051
21	θs	4.912	2.170	7.082
22	θs+n+α	4.910	2.185	7.095
23	α+θs+θr+n+Ks	4.911	1.962	6.873
improved soil				
1	α	4.911	2.484	7.395
2	Ks	4.922	2.402	7.324
3	Ks+α	4.914	2.329	7.243
4	Ks+l	4.926	2.673	7.600
5	Ks+θr	4.917	2.345	7.262
6	l	4.911	2.125	7.036
7	n	4.914	2.984	7.897
8	n+θs	4.911	2.207	7.118
9	θr	4.909	2.383	7.292
10	θr+θs	4.895	2.583	7.478
11	θs	4.906	2.878	7.784
potting mix+improved soil				
1	α+n	4.915	2.169	7.084
2	α+θs	4.918	2.044	6.962
3	n	4.906	1.618	6.524
4	Ks+l	4.911	2.132	7.043
MAX SCORE		**5**	**5**	**10**

Table 86: The columns report the scores of each indirect simulation in the first and second calibration phases. In the last column the final score (sum of the scores of the first and second calibration phases) for each simulation. In red the simulation with the highest final score.

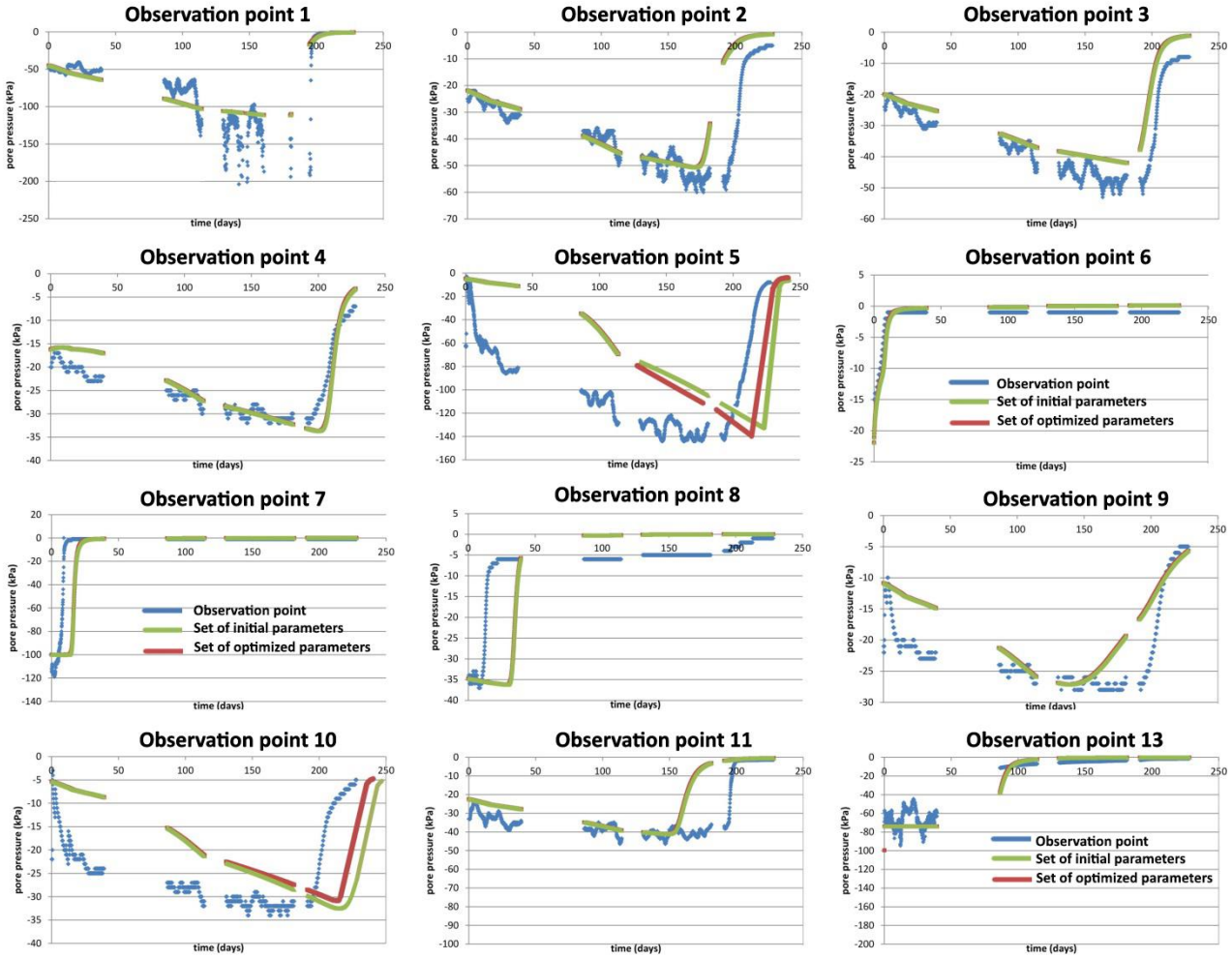


Figure 77: Evolution of suction with time using the set of initial parameters (green points), the set of optimized parameters (red points) and the observation points (blue points) in the observation nodes considered.

Table 87: Set of metrics/indices showing the enhancement of the model output in the observation points 5 and 10 using the set of optimized parameters instead of the set of initial parameters.

SET OF OPTIMIZED PARAMETERS										
	IA	**IA_j**	**IA'**	**NSE_j**	**NSE**	**R^2**	**MSE**	**MAE**	**RMSE**	**CRM**
Observation point 5	0.626	0.423	0.402	-0.196	-0.389	0.179	314.94	47.73	55.94	0.2069
Observation point 10	0.232	0.134	0.066	-0.869	-1.505	0.232	149.19	10.57	11.87	0.0611
SET OF INITIAL PARAMETERS										
	IA	**IA_j**	**IA'**	**NSE_j**	**NSE**	**R^2**	**MSE**	**MAE**	**RMSE**	**CRM**
Observation point 5	0.621	0.419	0.397	-0.206	-0.414	0.170	351.68	48.128	56.429	0.2186
Observation point 10	0.226	0.130	0.056	-0.888	-1.672	0.123	179.16	10.675	12.264	0.0982
Best fit	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**0**	**0**	**0**

Table 88:Set of initial parameters and set of optimized parameters of the investigated model. Indirect simulation 13.

SET OF INITIAL PARAMETERS						
	θr (-)	**θs (-)**	**α (1/m)**	**n (-)**	**Ks (m/sec)**	**l (-)**
Improved soil	0.0001	0.45	4.26	1.422	$2.93*10^{-7}$	0.5
Potting mix	0.0000	0.66	0.0052	1.453	$8.42*10^{-6}$	0.5
Gravel	0.0650	0.41	8.00	1.890	$1.22*10^{-5}$	0.5
SET OF OPTIMIZED PARAMETERS						
	θr (-)	**θs (-)**	**α (1/m)**	**n (-)**	**Ks (m/day)**	**l (-)**
Improved soil	0.0001	0.450	4.26	1.422	$2.93*10^{-7}$	0.5
Potting mix	0.0000	**1.757**	0.0052	1.453	$\mathbf{7.86*10^{-6}}$	0.5
Gravel	0,0650	0.410	8.00	1.890	$1.22*10^{-5}$	0.5

3_3_3 Preliminary conclusions and further studies

A sensitivity analysis has been performed prior to the calibration procedure in order to individuate the most relevant parameters for the analysis. The OFAT technique has been applied: the sensitivity coefficient has been calculated varying, one at a time, the investigated parameters of +/-1% of their initial value. For what concerns the parameters of the VGM, it has been observed a relevant sensitivity only at the arrival of the hydraulic stimulus and a decrease afterwards. The values of the sensitivity coefficient go to zero when the complete saturation is reached. The same was observed for the 1D filtration in column. It has been observed that the sensitivity coefficients obtained for the experiment in the soil column are greater compared to the ones observed for the large-scale apparatus experiment. This means that the 1D experiment is more suitable for the calibration of the hydraulic parameters. The sensitivity analysis has been applied to the parameters of the improved soil and potting mix. The α and θs parameters of the potting mix, which have sensitivity coefficients between 0 and -1, are the only ones to which the calibration programme should not apply to. The same procedure has been applied to the parameters of the water stress function of Feddes et al (1978). As expected, the observation points closer to the root bulb show the higher sensitivity to parameters variation. Moreover the observation points on the side of the bulb show an higher sensitivity with respect to the observation points below the bulb. With great probability this is due to the fact that the wetting front has an earlier influence on observation points below the root bulb with respect to the observation points located farther from the water table (the sensitivity coefficient decreases as the saturation increases). It has been observed that the higher influence on the model output is shown by the parameters h_{3high} and h_{3low} (values of pressure head below which roots can no longer extract water at the maximum rate), while T_{p1} and T_{p2} (potential transpiration rate) have no influence on the simulation output. For what concerns the sensitivity analysis of the parameters of the RWU spatial distribution function, as observed for the water stress function, the greatest sensibility is shown by the observation points on the side of the root bulb, with an increase of sensitivity passing from the more external point to the closer point. The parameters that have the highest sensitivity on the outputs are the maximum rooting lengthX_m, the depth of maximum intensity Z_m and with a lower relevance the depth of maximum intensity z^* and the radius of maximum intensity x^*. In contrast, the output seems to be little sensitive to the empirical parameter for vertical and horizontal direction, p_z and p_r.

The parameters of the RWU spatial distribution function of Vrugt et al (2001) are calibrated using a ‘’trial and error procedure’’ in order to obtain the best fit between observed-simulated datasets. The outcome is not satisfactory because Vrugt’s function imposes many constraints on the RWU pattern and for this reason

it is not possible to tune adequately the outcome of the model to the observation points in the observation nodes. Vrugt's function is not able to reconstruct the RWU spatial distribution with high accuracy. To overcome the limitations given by the adoption of a spatial function another technique has been chosen. Instead of using pre-determined functions, the dimensionless parameter is defined in the observation nodes using the information in time given by the sensors in order to minimize the differences between observed-simulated datasets then in the remaining points the values are obtained for interpolation. The calibration is operated considering the first 6 days of the simulation (after the last irrigation of the plant along the stem), because in this period the soil close to the bulb is saturated so the water stress is zero and as seen in the sensitivity analysis the influence of the hydraulic parameters on the simulation is as well zero in this period of time. A set of metrics/indices has been used to compare the performance of the model which uses the Vrugt's spatial distribution function and the model which uses the distribution elaborated by the modeler from experimental data. The metrics consistently show that the RWU spatial distribution elaborated using only the experimental data and any spatial distribution function better represents the experimental observations. To calibrate the compensatory RWU mechanism, the ''trial and error'' procedure has been used. In particular the critical water stress index has been incrementally increased from 0.6 to 1. It has been observed that the influence of the compensatory mechanism increases approaching the root bulb and it is maximum in the observation points on the side of the bulb (compared to the observation points below). A set of metrics has been used to identify the best value of the water stress index which assures the lowest residual between observed-simulated datasets. A critical stress index of 1 (which means zero compensatory ability of the plant) shows the best match. In fact in accordance with Cai et al (2017) and Simunek and Hopmans (2009), cultural plants have a low (or zero) ability to compensate for water stress. A calibration of the hydraulic parameters of the potting mix and improved soil has been performed applying the same simplified procedure in two phases adopted for the 1D filtration in the soil column. The first phase uses metrics/indices to evaluate the match between observed/simulated datasets, while the second phase focuses on the results of the inverse analysis and the set of the optimized parameters individuated. As performed also in PART I of the present thesis, different combinations of parameters have been optimized in the indirect problem belonging to the single layer and both the layers (improved soil+potting mix). The indirect simulation that shows the higher score optimizes the hydraulic parameters Ks and θs of the potting mix. An enhancement in the model performance is observed only for the observation points closer to the root bulb (distance of 30 cm in the horizontal and vertical direction). The scarce enhancement achieved is due, with high probability, to the low sensitivity of the hydraulic parameters in the investigated problem as observed in the sensitivity analysis. In fact the described phenomenon is affected mostly by the RWU spatial parameters and by the evapo-transpirative phenomenon. On the contrary, the 1D filtration experiment is highly influenced by SWRC parameters and in this case, inverse analysis, is optimal to perform parameters optimization. Moreover the enhancement concentrated only in the potting mix layer makes us reflect on the limits of the use of inverse analysis when dealing with more than one single layer. In fact in multi-layered domains, the enhancements of model performance are often localized in one single material and rarely groups of materials give the greatest enhancement among the performed indirect simulations. This is due to the fact that problems of uncertainty, non-identifiability and correlation of the optimized parameters increase increasing the number of parameters to optimize.

In the next chapter 3_4, the final considerations on the output of the elaborated model after the performed in-depth sensitivity analysis and parameters calibration are reported extensively. Subsequently

the equipment of the large-scale apparatus together with the obtained laboratory results in the different experimental phases are reported and discussed.

3_4 SOME CONSIDERATIONS ON THE OPTIMIZED MODEL

In Figure 78 the time lapse of the progressive pressure head distribution in the large scale apparatus as the wetting front propagates upwards is presented. Initially the root bulb depletes of water the closer soil volume as could be observed comparing the pressure head distribution of day 0 to the one of day 24,5 in close proximity of the bulb. Subsequently a gradient occurs from the wetter external part of the soil volume toward the root bulb which starts depleting the water of the farer soil volume. The soil volume on both sides of the root bulb passes from a pressure head of 0 kPa at day 0 to a pressure head of -120 at day 190. In the meantime, fine roots start growing toward the farer zone of the domain richer in water and nutrients. As the saturation profile moves upward, the RWU increases in the wetter part of the soil volume located below the root bulb with respect to the zones on the sides of the bulb which are in drier conditions. On the contrary the soil volume above the root bulb is rapidly depleted of water due to the combination of evaporation and transpiration phenomena.

For what concerns the evaporation from the soil surface it is close to zero in phase 1A and it increases in phase 1B due to a reduction of the relative humidity from ~99% to 70%. A peak of evaporation outflow is observed in phase 1B then the creation of a crust of very dry soil on the soil surface together with a reduction in the saturated permeability (from 0.00020 m/day to 1-2*10^{-6} m/day) reduces even more the exchange with the atmosphere. After that, the outflow due to evaporation is pretty constant in time because the surface is inside an enclosed space, out of direct sunlight so there is no cyclic behaviour due to alternation of day and night.

The inlet cumulative water flux from the lower boundary (variable pressure head boundary condition) is 0.18 m^2 [6] ,the cumulative outflow flux due to evaporation is equal to 0.19% (0.00035 m^2); the cumulative flux of the potential root water uptake is equal to 11.11% (0.020 m^2) while the cumulative flux of the actual RWU is equal to 7.22% (0.013 m^2).

[6] In Hydrus the cumulative water flux for a 2D model (dimensions X,Z) is computed with dimensions $L^2 = L^3/$(L in the perpendicular direction (Y)).

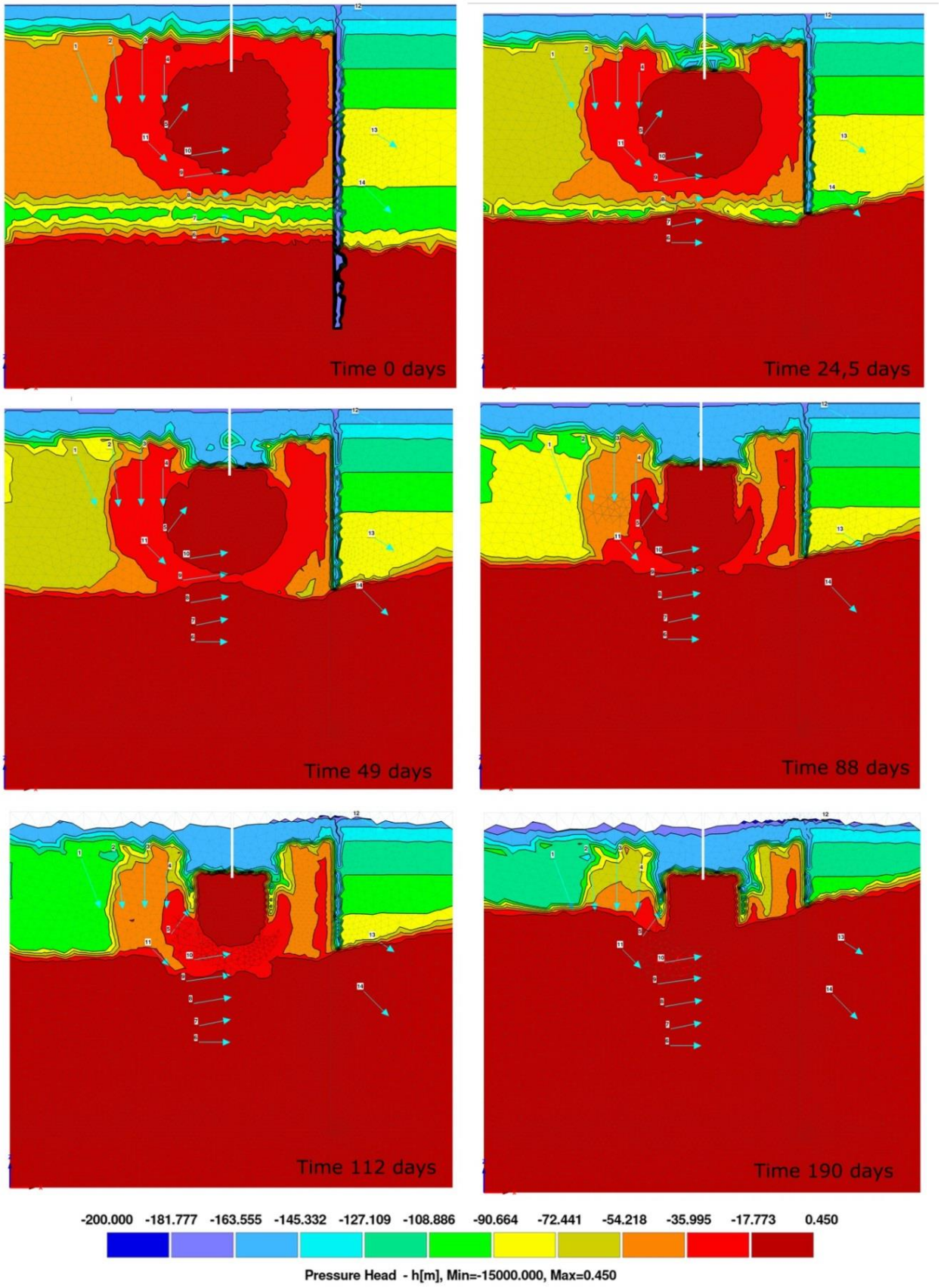


Figure 78: Time lapse of the progressive pressure head distribution in the large-scale apparatus as the wetting front propagates upwards. The effects of the RWU of the plant on the pore pressure distribution are clearly observables.

3_5 DESCRIPTION OF THE LARGE-SCALE EXPERIMENTAL SET UP

A specific apparatus has been developed to investigate the spatial and temporal evolution of suction in relation to root water uptake of Malaleuca Styphelioides. In the adult stage, this evergreen native plant could achieve 20 m of height. It is tolerant to drought, wet soil and smog. It is suitable for full sun exposure and it has a low water need. The plants used for the project are 40cm in height and they have been grown in a potting mix extremely organic.

A sketch of the apparatus is presented in Figure 79.

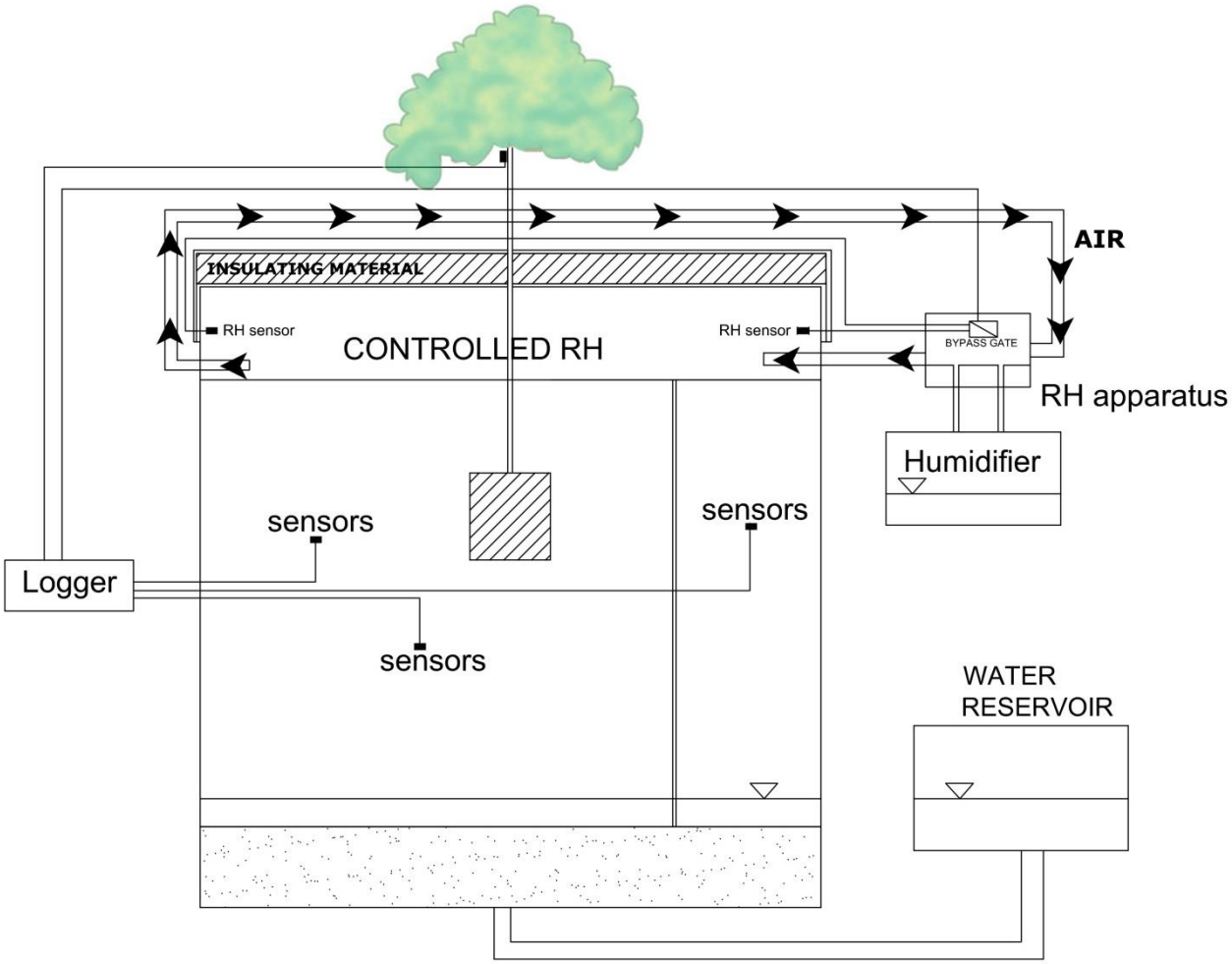


Figure 79: Sketch of the large scale apparatus designed and built in laboratory

The setup is composed of a plastic container (volume of 1 m^3), filled with 720 mm of improved soil (35% fine material, 65% sand) overlying 130 mm of clean gravel. A non-woven fabric was used as a separation layer between the gravel and the improved soil. The improved soil was placed at an initial water content of 10% and in layers of 50 mm, achieving an average void ratio of 0.8.

The plant was positioned exactly at the centre of the container with about 150 mm of soil cover to reduce the effect of evaporation. The dimension of the root bulb is equal to the dimension of the pot in which the plant has been grown that was completely filled with roots at the time of the plant installation. The bottom

of the root bulb was at an approximate depth of 290 mm from the soil surface. There is hence 430 mm between the underside of the root bulb and the top of gravel. A small tube placed along the plant stem was used to water the plant from the top to ensure the survival of the plant just after the replant and in the transient period between the application of the water level and the instauration of a steady-state seepage. It was not possible to separate the roots bulb from the potting mix in which the plant has been grown because the mechanical separation of the roots from the soil causes the breakage of the smallest roots, through which absorption of water and nutrients occurs.

The water level at the bottom of the container was established via a water reservoir connected to the base of the apparatus. The water level was maintained constant by means of a float valve at the water reservoir inlet. The water reservoir was covered with a plastic sheet to reduce evaporation in the reservoir. During the test, some fertilizer was poured in the water reservoir to provide nutrients to the plant.

The top of the container was closed by a plastic cover and some insulating panels in order to create an environment where the relative humidity (RH) can be controlled. This was achieved via a specifically designed apparatus that can circulate some air (via a pump) through a water reservoir and through the volume where the RH is to be controlled. Two RH sensors are used to measure the RH above the soil surface and the flow of air is adjusted until the target RH is achieved. Note that there is no temperature control in the experiment, the insulating panels are simply used to reduce daily temperature oscillations which affect the suction above the soil surface (as per psychrometric law).

A grow light (model Hi-Par Dynamic CMH 315w DE Control Kit) was placed above the plant to ensure a similar light spectrum to natural sunshine (full spectrum). The photosynthetic photon flux density (PPFD) of the chosen grow light is 856.61 µmol $m^2 \cdot s^{-1}$ tested by the manufacturer at a height of 60 cm from the soil surface, which is equal to 50539.99lux (186,741 W/m^2), the conversion has been made according to the type of lamp (ceramic metal halide). The PPFD measures the number of photosynthetically active photons that fall on a given surface each second, which is an objective measure since it directly indicates how much light energy is available for plants to use in photosynthesis. The lamp was positioned at 60 cm from the top of the plant as suggested by the producer and 10 to 12 h of light were provided per day.

Teros 21 sensors (Meter group) were used to monitor the evolution of suction in time. These sensors indirectly measure suction by estimating the dielectric permittivity of two ceramic discs in hydraulic equilibrium with the surrounding soil. The dielectric permittivity is then converted to volumetric moisture content and to suction via the moisture characteristic curve of the ceramic discs. The sensors have a working range from -9 kPa to -100,000 kPa with an accuracy of (+/- 10% of the reading +2kPa) as indicated by the manufacturer. Each sensor was calibrated by the producer with a five-point calibration curve.

The Teros 21 sensors were preferred to tensiometers, which offer a direct measurement of suction, because they do not suffer from cavitation, a phenomenon plaguing tensiometers. Their main limitation is the inability to measure pressures greater than -9kPa. Teros 21 sensors also measure soil temperature with an accuracy of +/-1°C.

The location of all suction sensors is given in cross sections of Figure 80. A corner of the box is isolated from the rest of the soil body by an impervious boundary and serves as a control zone (top right corner of the plane views in Figure 80).

Three sensors were placed in the control zone (cross-section BB) at depths of 75 mm, 320 mm and 470 mm from the soil surface. Five sensors were installed under the centre of the root bulb (see cross-section AA in Figure 80) at different depths to investigate changes in suction with depth under the root bulb. The shallowest sensor is 30 mm below the underside of the bulb and the other sensors are 50 mm apart. Then, five sensors were installed at different radial distance of the side of the root bulb and at the depth of 220 mm from the soil surface. The closest sensor is at 30 mm from the bulb and the other sensors were placed 50 mm apart. Finally, one sensor was installed at 320 mm below the underside of the root bulb and at a radial distance of 145 mm from the centre of the bulb.

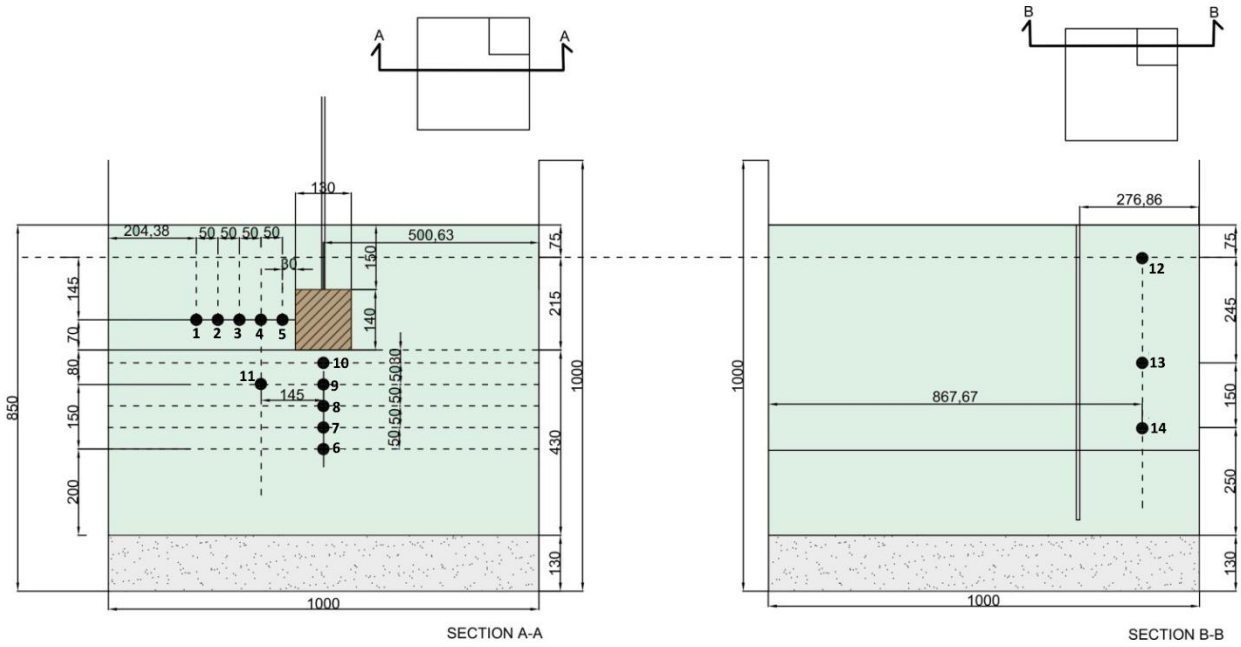


Figure 80: Cross sections of the large-scale apparatus with indications of the exact positions of the sensors with respect to the sides of the box and the root bulb. All dimensions are in mm.

Figure 81 shows photos of the apparatus being set up with the barrels of improved soil ready to be used (A); the non-woven fabric covering the gravel on the bottom of the box and the first layer of the soil above it (B), installation of one sensor below the root bulb (C); installation of sensors on the side of the root bulb (D); installation of a sensor in the control zone with view of the steel sheet used to isolate the control zone (E); soil around the plant being wet in order to remove the plastic pot and insert the root bulb without collapse of the hole (F); plant after plastic pot removal (G) and box completely filled with soil (H).

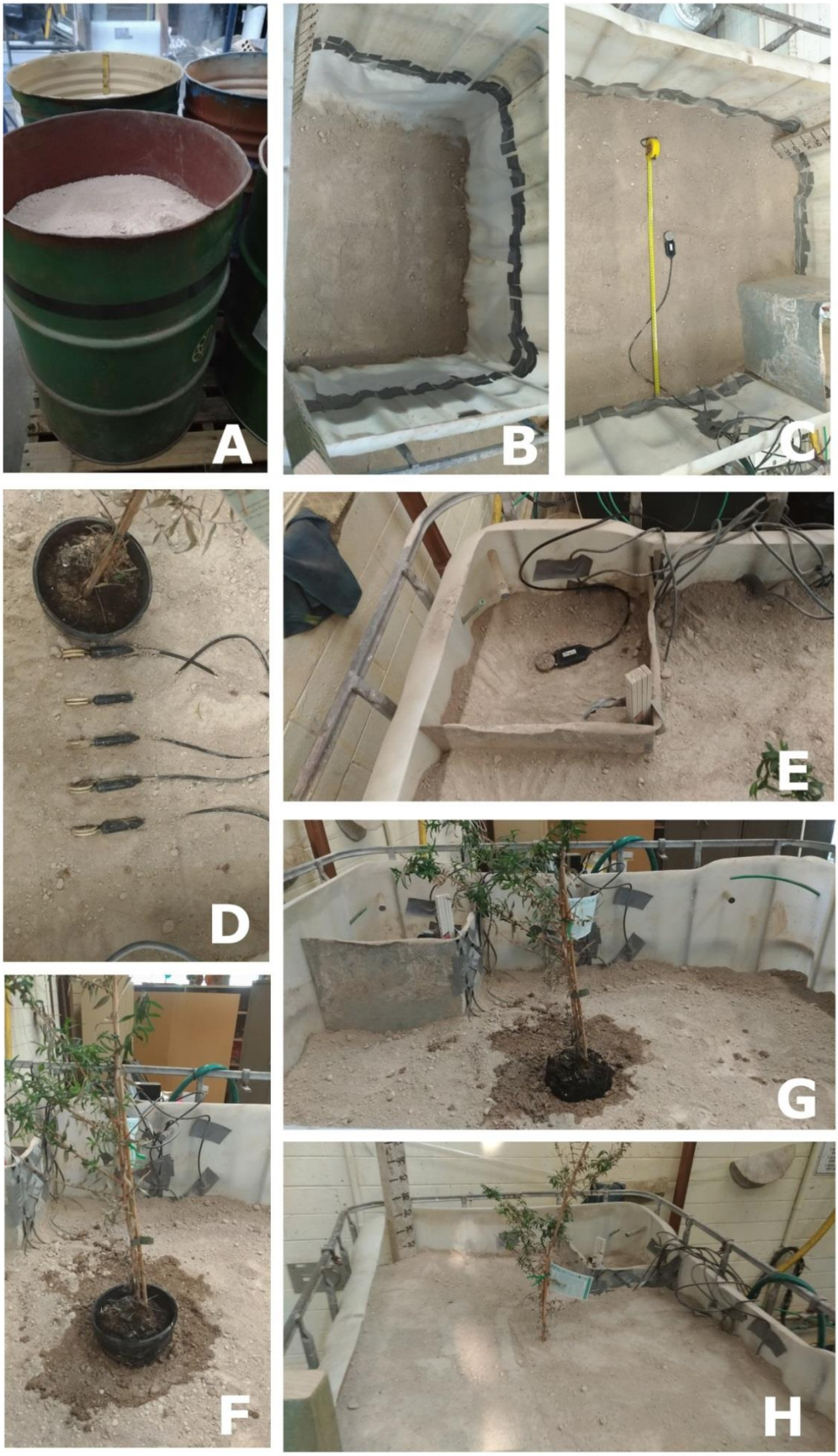


Figure 81: Some photos of the large-scale apparatus during the set-up

3_6 EXPERIMENTAL PROGRAM

The water level and the relative humidity were changed during the experiment to track the spatial and temporal evolution of suction around the root bulb and provide insight into the root water uptake (RWU). Table 89 provides a summary of the different phases of testing and boundary conditions applied.

Table 89: Time schedule of the different phases of the large-scale apparatus testing.

	RH [%] top	Water level [mm] from the base of the apparatus	Water level [mm] from the soil surface	Time	Time (day)
Phase 0	uncontrolled	135 at day 15	715 at day 15	From 4/12/2019 to 5/1/2020	0 – 30
Phase 1A	~99%	200	650	From 5/1/2020 to 7/4/2020	30 – 117
Phase 1B	~74%	200	650	From 7/4/20 to 12/5/2020	117 - 152
Phase 2	~78%*	355	495	From 12/5/2020 to 20/7/2020	152 - 220
Phase 3	~78%*	432	418	From 20/7/2020 to 4/9/2020	220 - 266
Phase 4	~99%	312	538	From 4/9/2020 to 16/10/2020.	266-308

* note that RH was not deliberately changed from 74% to 78% between phase 1B and 2. The value of RH reported is actually an average measured value (not the target one), which is affected by the RH in the building where the experiment was conducted, despite RH control system.

Details about the different phases are given below:

- Phase 0 (duration ~ 1 month) corresponds to an equilibration phase right after setting up the experiment and planting the Melaleuca seedling. In this phase, the plant was kept alive by irrigation through the tube along the plant stem. Irrigation was essential since the box was filled with dry soil (below wilting point). the water level was set at 135 mm from the bottom of the box (5 mm above the gravelly layer) at day 15.
- In phase 1A (duration ~ 3 months), the relative humidity above soil surface was kept close to 100% to avoid evaporation and the water level was set at 200 mm from the bottom of the box (70 mm above the gravelly layer).
- In phase 1B (duration ~ 1 month), the water level was kept at 200 mm but the relative humidity above soil surface was dropped to 75% in order to observe the combined effect of evaporation and transpiration.
- In phase 2 (duration ~ 2 months), the water level was raised to 355mm from the bottom of the box (with a relative humidity close to that of phase 1B) in order to observe the root bulb behaviour in a condition close to soil saturation, the same condition that occurs due to the raising of the shallow water level in the rainiest months.
- In phase 3 (duration ~ 2.5 months), the water level is further raised to 432 mm, simulating a period of high water level for the plant, again evaporation and transpiration are investigated together.
- In phase 4 (duration ~ 1.0 months), the water level was dropped to 312mm and the relative humidity was raised to 99% in order to reduce evaporation and assess the effect of RWU on a drying path.

The boundary conditions are applied both to the control zone (without roots) and in the main part of the experiment, i.e. containing the plant.

3_7 EXPERIMENTAL EVOLUTION OF SUCTION IN TIME

In this section, the whole evolution of suction recorded by all sensors is first presented and then, each phase is discussed one by one. As a matter of convention, suctions are reported as positive values while pore pressure as negative values: a suction of 500 kPa corresponds to a pore pressure of -500kPa.

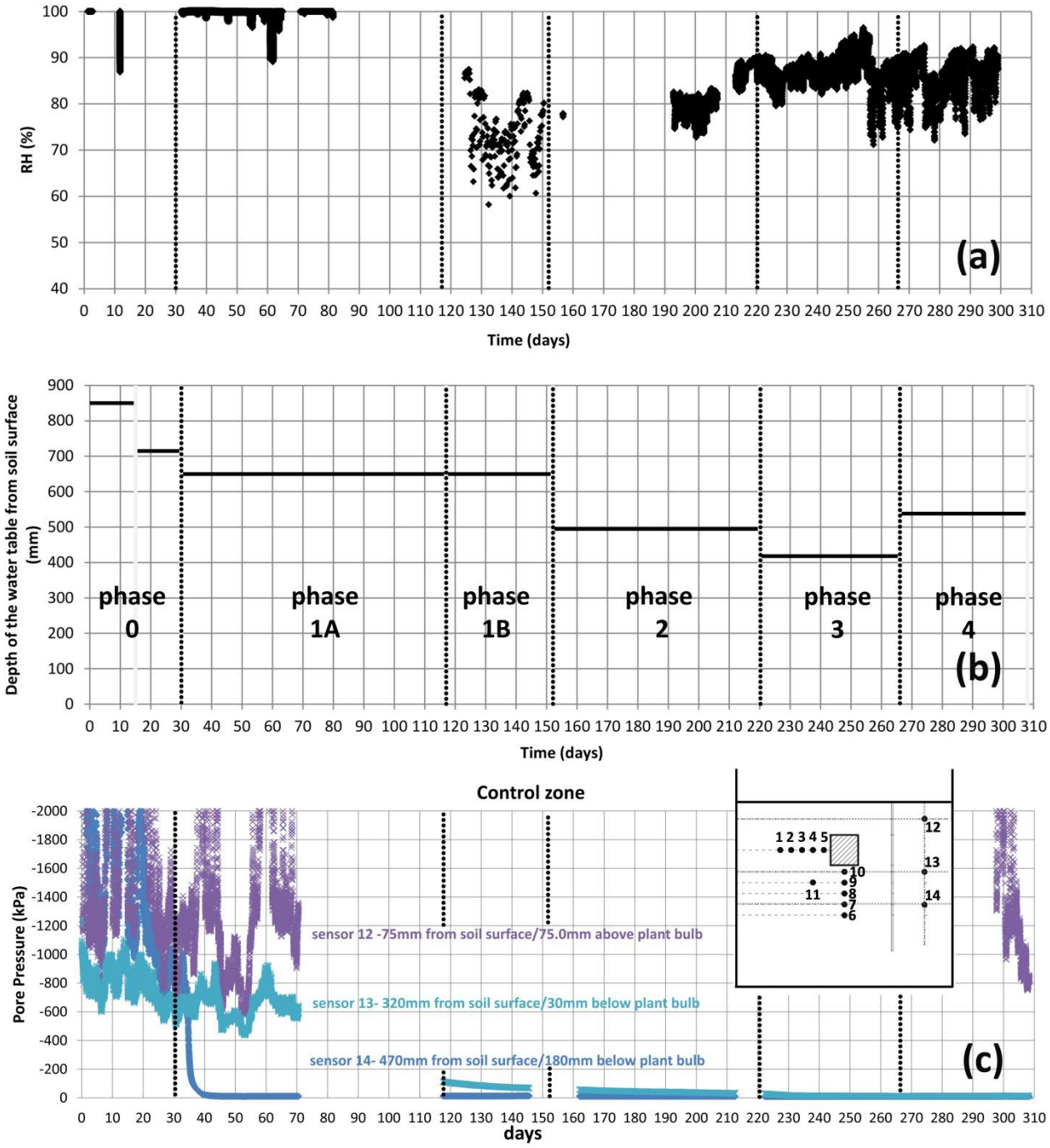


Figure 82: The first graph reports the Relative Humidity in the upper boundary, the second graph the depth from the soil surface of the imposed water table. The pore pressure values in time recorded by the installed sensors in the control zone (12, 13, 14) in the different testing phases are reported in the third graph.

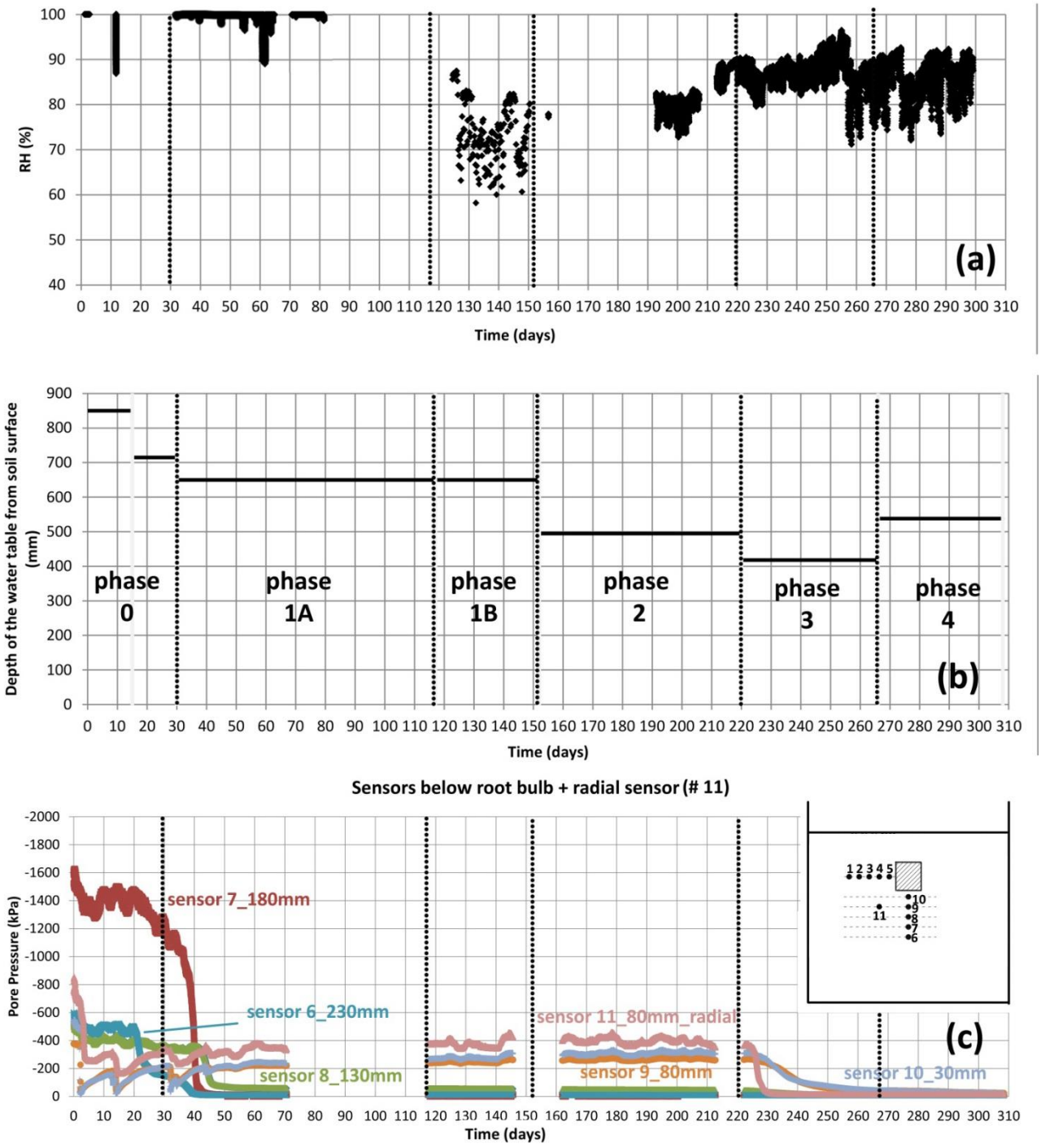


Figure 83: The first graph reports the Relative Humidity in the upper boundary, the second graph the depth from the soil surface of the imposed water table. The pore pressure values in time recorded by the installed sensors below the root bulb (6, 7, 8, 9, 10, 11) in the different testing phases are reported in the third graph.

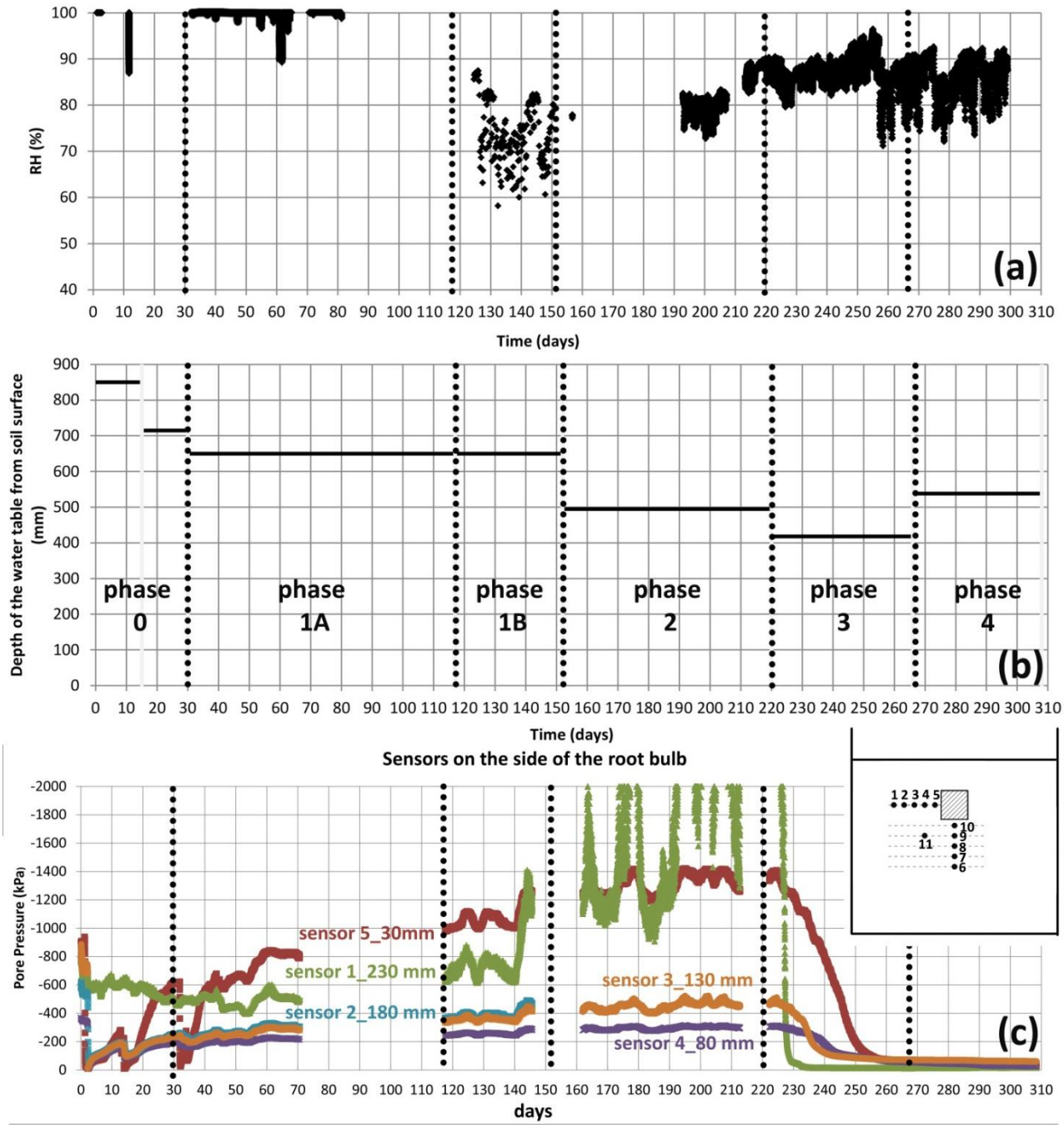
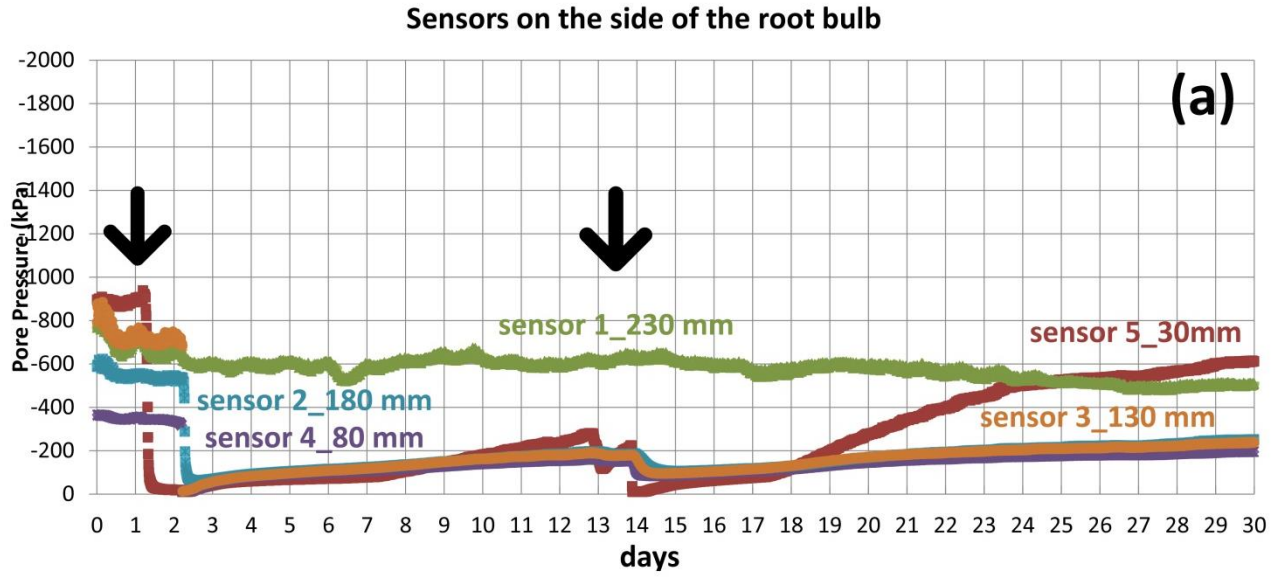
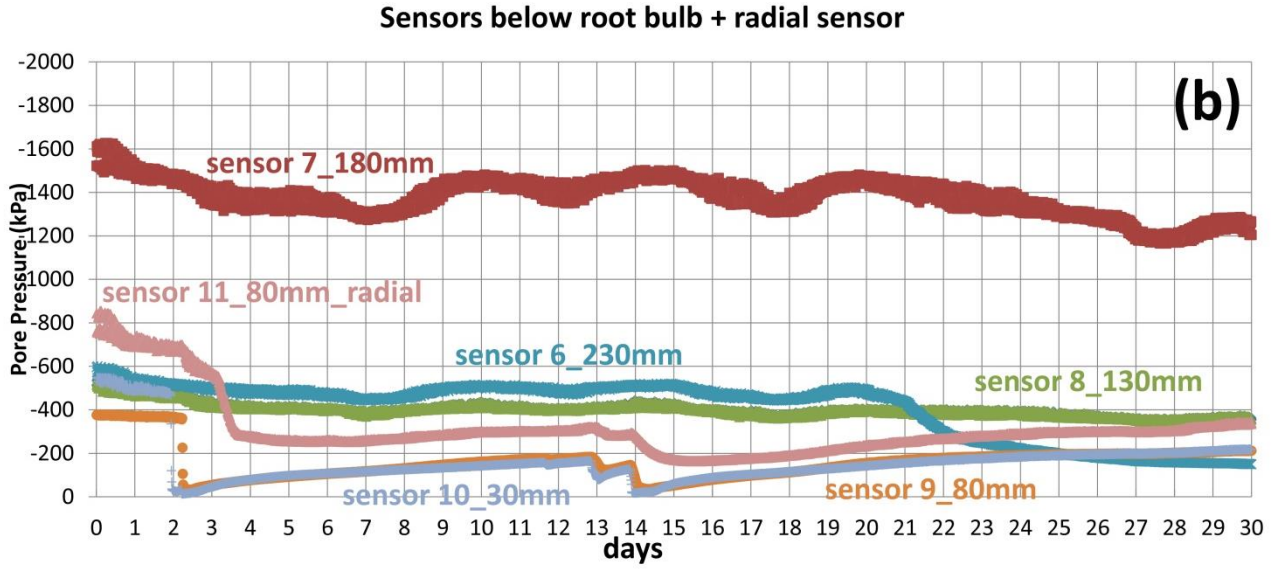


Figure 84: The first graph reports the Relative Humidity in the upper boundary, the second graph the depth from the soil surface of the imposed water table. The pore pressure values in time recorded by the installed sensors on the side of the root bulb (1, 2, 3, 4, 5) in the different testing phases are reported in the third graph.

3_7_1 Phase 0

The data pertaining to phase 0 are shown in Figure 85 and the results are discussed zone by zone (control zone, on the side of the bulb, below the bulb). This is the first phase of the experiment where the soil is initially at a very high suction (800-900 kPa) and water is brought in the box for the first time. As such, all initial values of suction read by the sensors are very high, fluctuating and progressively come to equilibrium.


Sensors on the side of the root bulb
(a)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 1_230 mm
sensor 5_30mm
sensor 2_180 mm
sensor 3_130 mm
sensor 4_80 mm
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
days


Sensors below root bulb + radial sensor
(b)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 7_180mm
sensor 11_80mm_radial
sensor 6_230mm
sensor 8_130mm
sensor 10_30mm
sensor 9_80mm
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
days

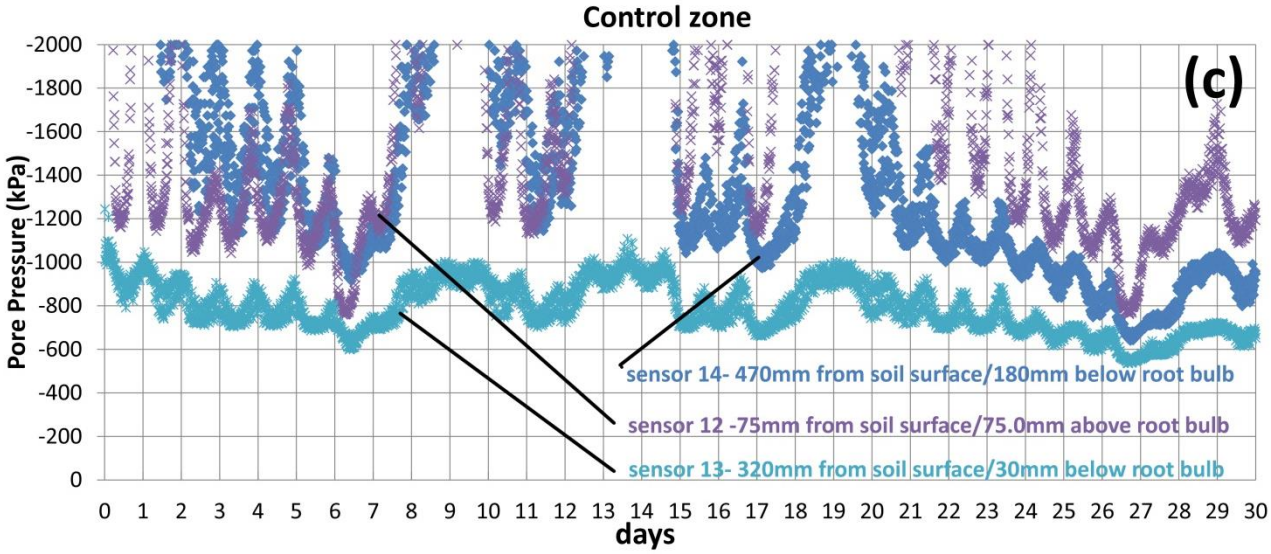

Control zone
(c)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 14- 470mm from soil surface/180mm below root bulb
sensor 12 -75mm from soil surface/75.0mm above root bulb
sensor 13- 320mm from soil surface/30mm below root bulb
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
days

Figure 85: Evolution of pore pressure with time on the side of the root bulb (a), under the root bulb (b) and in the control zone (c). Water level is set to 715 mm at day 15 from the soil surface. RH is uncontrolled (the top surface of the soil is not closed with a cover in this phase).The black arrows indicate the time instants when irrigation along the plant stem has been carried out.

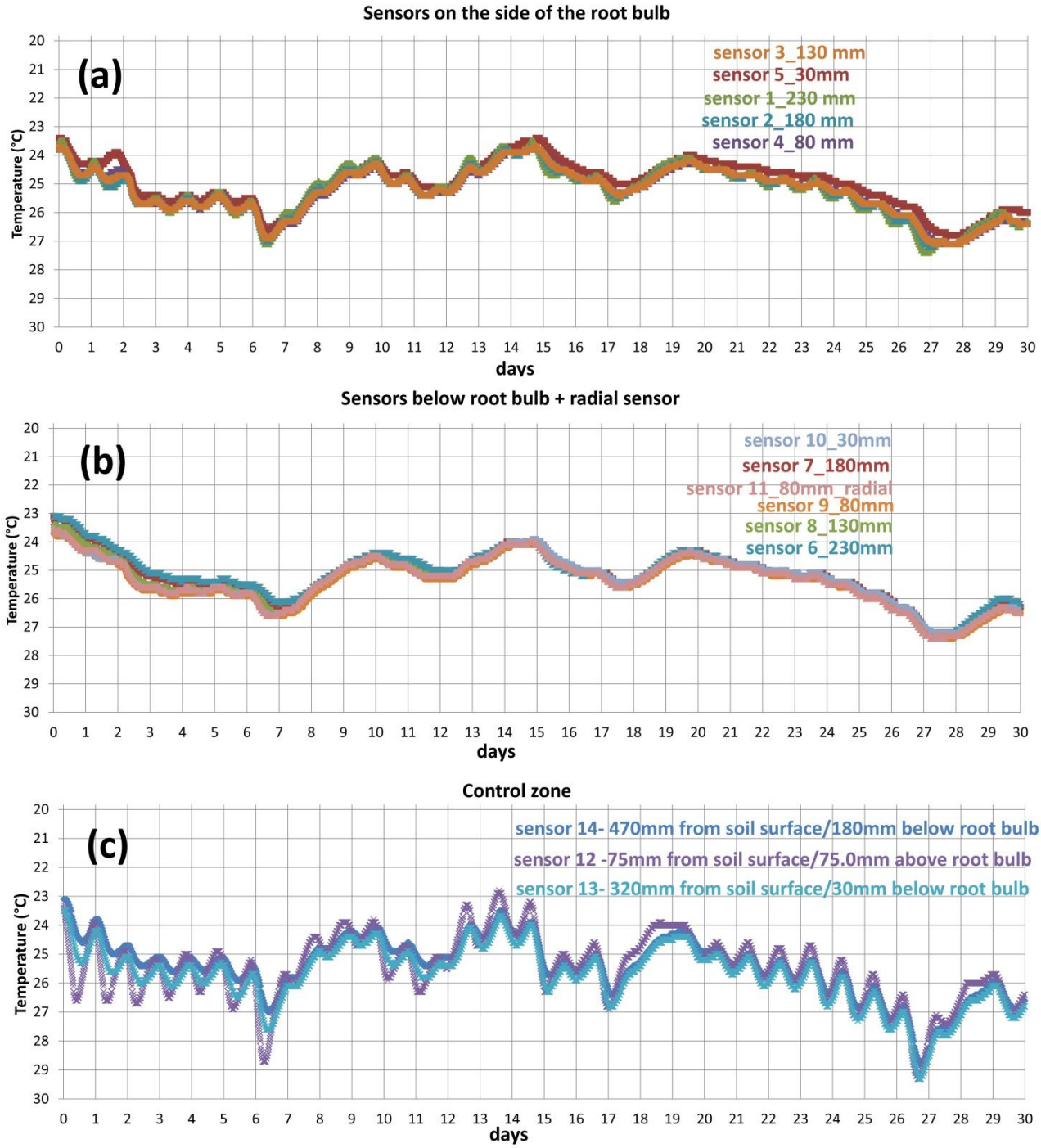


Figure 86: Evolution of temperature with time on the side of the root bulb (a), below the root bulb (b) and in the control zone (c). Water level is set to 715 mm at day 15 from the soil surface. RH is uncontrolled (the top surface of the soil is not closed with a cover in this phase).

3_7_1_1 Control Zone

Significant fluctuations are recorded in the control zone in the initial phase of the experiment. First the sensors were air dried when installed in a dry soil, so the suction values were either out of range or very

high. At the high end of suction, the sensors are quite sensitive to temperature fluctuation. Figure 86 shows the fluctuation of temperature recorded by all 3 sensors (depth of 75mm, 320mm, 470 mm) with less fluctuation amplitude as the sensor is deeper in the soil (compare the temperature trend of sensor 12 and sensor 14), which is expected given the extra insulation provided by thicker soil cover. No meaningful conclusion can be drawn during this phase for the control zone.

3_7_1_2 On the side of the root bulb

Figure 85(a) shows the temporal evolution of suction at a radial distance from the root bulb (indicated on the Figure). To interpret the sensors' response, it is important to consider that water was provided to the plant at 2, 14 and 32 days (from the tube positioned along the stem of the plant creating a sphere of wet soil around the plant) and that the plant starts taking water from the moment it is re-planted.

Sensor 5 is the closest to the plant and shows strong fluctuations: suction drops following wetting and then progressively increases as the plant uptakes water. After wetting at day 14, it can be seen that the plant generates a suction of 600 kPa in 15 days, which is quite a significant value of suction (~halfway to the wilting point). Similar trends, but with lesser amplitude, can be observed for sensors 4, 3, 2. A suction increase of about 95 kPa is generated by the plant between days 14 and 31 in observation point 4 (80 mm from the root bulb) and 85 kPa in observation point 3 (130 mm from the root bulb).

In contrast, sensor 1 is too far (230 mm) from the root bulb to react to wetting events and root water uptake. Its response corresponds to a slow equilibration with the surrounding soil and, with 220 mm of cover, no temperature influence is expected (as confirmed by its record of temperature, see Figure 86a). Figure 87 shows the increment in suction observed for the sensors on the side of the root bulb (graph a) and below the root bulb (graph b) in the period 14-31 days. As it is possible to observe in graph b decrements of suction (increases in pwp) are observed for the deeper sensors (from the soil surface) due to the application of a water table (depth 715 mm from the soil surface) at day 15.

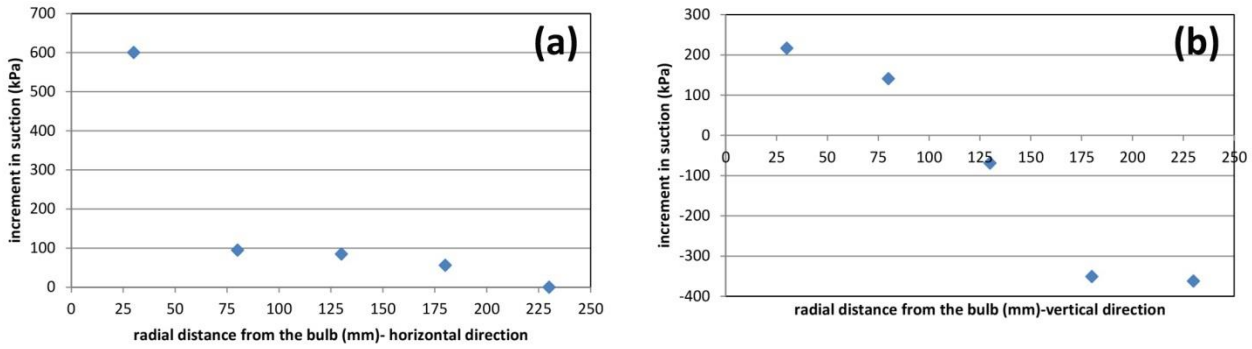


Figure 87: Increment in suction recorded from the sensors on the side of the root bulb (graph a) and below the root bulb (graph b) in the period 14 -31 days.

3_7_1_3 Below the root bulb

Figure 85 (b) shows the temporal evolution of suction for the sensors located under the centre of the root bulb and the offset sensor (#11). The first observation is that sensor 7 returns a very high initial value of suction (over 1000 kPa), compared to all other sensors (initial reading between 400 and 600kPa). This variability is due to the lack of control on initial condition of sensors and the soil being initially at a low water content. Again, phase 0 is an equilibration phase in the soil.

Like on the side of the bulb, a clear effect of the wetting events is visible in the sensor response, especially those closest to the bulb (e.g. sensors 9 and 10, as opposed to sensors 8 and 6).

In phase 0, water was introduced in the bottom of the box and the deepest sensor (6, 230 mm below the bulb) is the quickest to react to the wetting front, with a clear reduction in suction at about 20 days. Then sensors 7 (180 mm below the bulb) and 8 (130 mm below the bulb) react. It can also be observed that sensor 11, which is 80 mm below the bulb but with 145 mm of radial offset, follows a trend very similar to sensor 6 (same depth, under the centre of the bulb) but with higher suction, which makes sense.

Comparing the response of sensors 10 (30 mm under the bulb) and 5 (30 mm beside the bulb), it can be seen that the suction below the bulb is lower than the suction on the side of the bulb which can have different explanation: a possible difference in root density (unverified hypothesis) and/or a gravity effect where water is more likely to pond at the bottom of the very permeable potting mix and reduce suction preferably under the bulb, as opposed to the side.

3_7_2 Phase 1

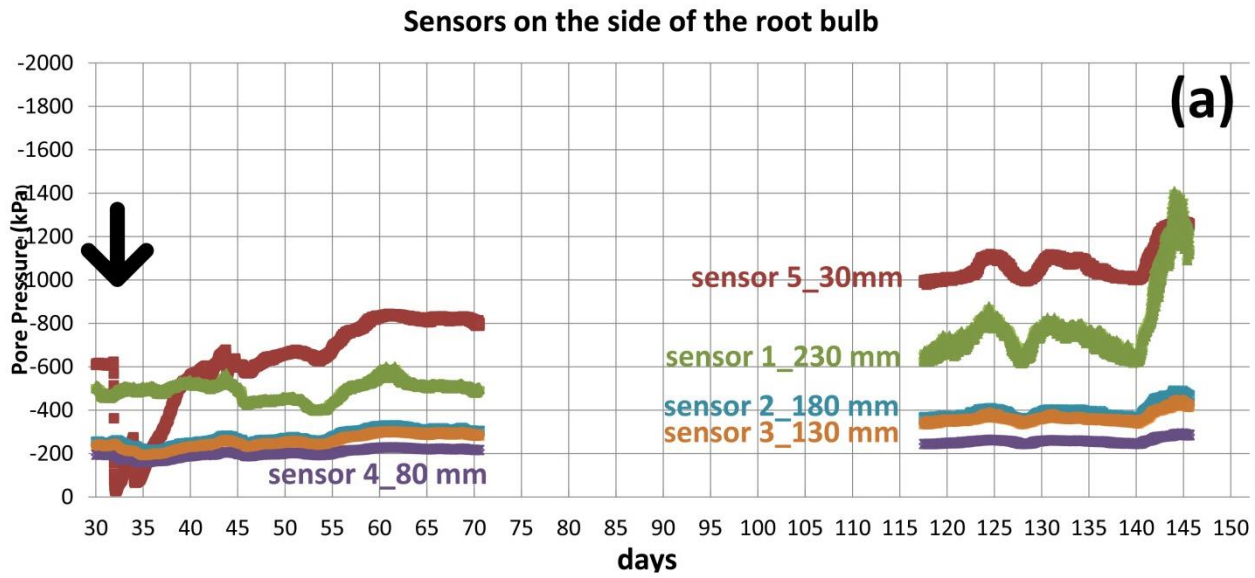
The data pertaining to phases 1A and 1B are shown in Figure 88 and the results are discussed zone by zone (control zone, on the side of the bulb, below the bulb). Please note that there was a logging issue between days 70 and 117, so not suction or temperature was recorded over that period.

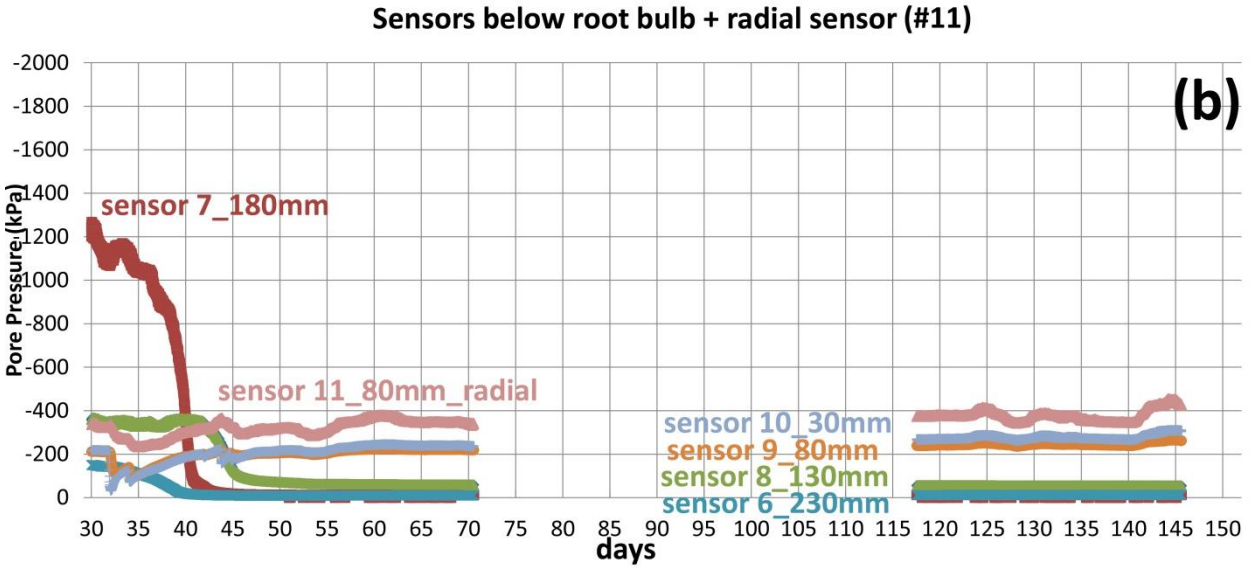
3_7_2_1 Control Zone

In phase 1A the water level is at 650 mm from the soil surface, generating an upward wetting front due to a suction gradient between the bottom and the top of the box. Sensor 6, being the deepest in the control zone (depth of 470 mm) is the quickest to respond: in less than 10 days (from days 33 to 42), suction drops from 1000 kPa to an approximatively constant value of 11 kPa. At that depth, there is little effect of evaporation from the surface. Sensor 13 (320 mm deep) also ends up at a low value of suction (68 kPa) although equilibration is slower (due to the slow process of upwards moisture migration). Note that the actual trend was not captured because of the logging problem. Being very close to the surface (75mm below it), sensor 12 is affected by the evaporation through the surface and it fell out of its measurement range sometimes during phase 1A (its value is meaningless and hence not reported).

The relative humidity was dropped from 99% to 74% between phase 1A and 1B (at about 117 days), resulting in an increased suction gradient. Although the trend of sensor 13 is not captured (because of the logging issue), an increase in the suction gradient should result in a higher value of suction. The fact that suction keeps reducing is an indication that equilibrium under the conditions of phase 1A was not attained.

Phase 2 started at day 152 before sensor reached equilibrium because suction close to the root bulb was approaching the wilting point (~1500 kPa, see next section) and the plant was at risk of significant water stress. Sensor 13 shows that the moisture movement in the large container is quite slow and several months may be required to reach steady state.


Sensors on the side of the root bulb
(a)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 5_30mm
sensor 1_230 mm
sensor 2_180 mm
sensor 3_130 mm
sensor 4_80 mm
30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150
days


Sensors below root bulb + radial sensor (#11)
(b)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 7_180mm
sensor 11_80mm_radial
sensor 10_30mm
sensor 9_80mm
sensor 8_130mm
sensor 6_230mm
30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150
days

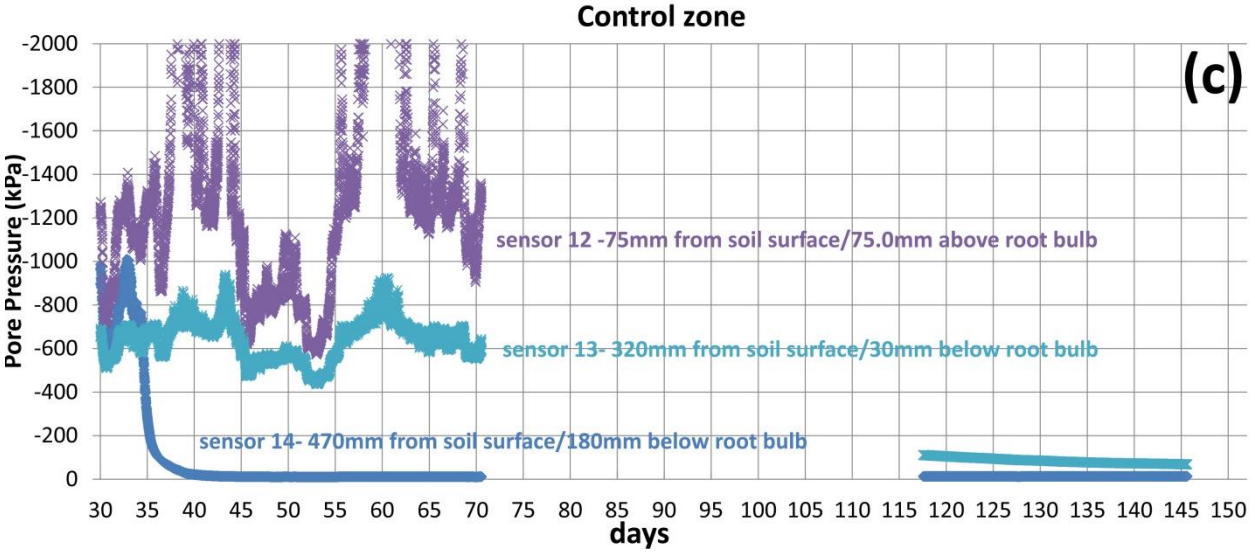

Control zone
(c)
Pore Pressure (kPa)
-2000
-1800
-1600
-1400
-1200
-1000
-800
-600
-400
-200
0
sensor 12 -75mm from soil surface/75.0mm above root bulb
sensor 13- 320mm from soil surface/30mm below root bulb
sensor 14- 470mm from soil surface/180mm below root bulb
30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150
days

Figure 88: Evolution of pore pressure with time on the side of the root bulb (a), under the root bulb (b) and in the control zone (c) in phase 1 (1A+1B). Water level is set to 650 mm from the soil surface in both the sub-phases. RH is set to~99% in sub-phase 1A and ~74% in sub-phase 1B. The black arrow ndicates the last irrigation along the plant stem.

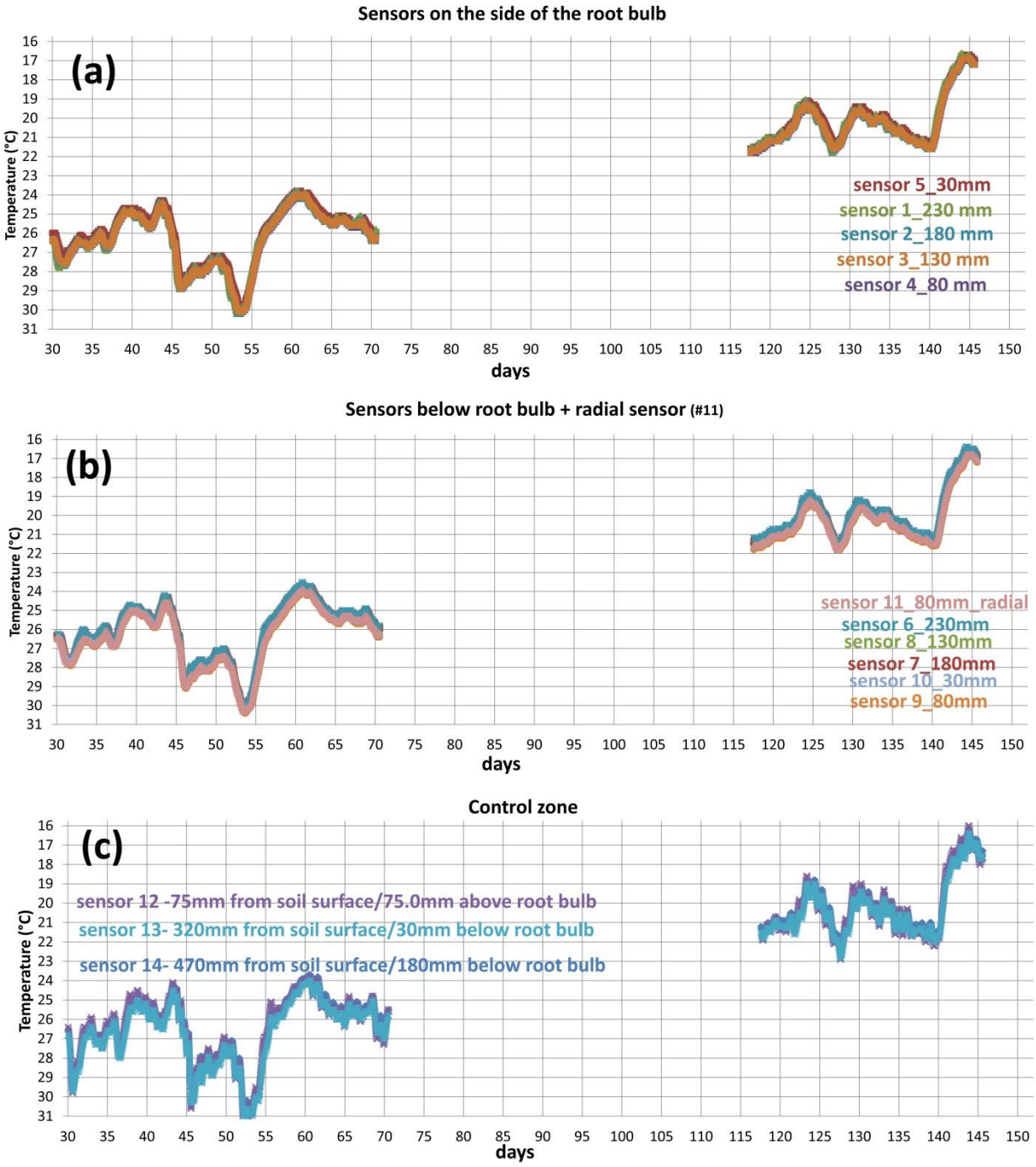


Figure 89: Evolution of temperature with time on the side of the root bulb (a), under the root bulb (b) and in the control zone (c) in phase 1 (1A+1B). Water level is set to 650 mm from the soil surface in both the sub-phases. RH is set to~99% in sub-phase 1A and ~74% in sub-phase 1B.

3_7_2_2 On the side of the root bulb

In phase 1, the RH was initially at 99% to avoid evaporation (phase 1A) and then dropped to 74% (phase 1B). All sensors discussed here are at the same depth. Three responses can be observed:

- Sensor 1 is the furthest from the bulb (230mm) and its suction reading seems to be solely affected by the surrounding soil, not the root uptake. Fluctuations of suction are due to fluctuations of temperature (which are the same for all the considered sensors at the same depth as could be observed in Figure 89).
- Sensor 5 is the closest to the root bulb and shows the highest value of suction because of a strong effect of root water uptake.
- Sensors 2, 3, 4 fall in between: they are within the zone of influence of the root bulb but far enough that suction does not increase as rapidly as for sensor 5. Apart from sensor 5, suction seems to be slightly increasing with radial distance, from sensor 4 to 1, which is the consequence of the irrigation events which have led to a radial wetting front. Then, water uptake has locally increased suction close to the bulb (sensor 5).

Overall, similar trends are visible during phase 1B. However, several points are worth noting:

- the increase in suction recorded by sensor 1 (of about 100 kPa) between the end of phase 1A and the beginning of 1B could be due to evaporation through the surface. However, sensors 2, 3 and 4 are not affected by the same amount, albeit under the same amount of cover, which is unexplained.
- the increase in suction recorded by sensor 5 between the end of phase 1A and the beginning of 1B is higher than for sensor 1 (about 200 kPa) could be due to a combination of evaporation through the surface and root water uptake.
- at day 140, all the sensors on the side of the root bulb display an increase in suction, particularly sharp for sensors 1 and 5. We observe also a fall in temperature from 21 to 16.9 degrees and an oscillation of the RH between 60-75%. The soil total suction ψ, which is the equivalent suction derived from the vapour pressure p_1 in equilibrium with the soil water, relative to saturation vapour pressure p_{o_1}, is related to temperature T and RH according to the psychrometric law (Fredlund and Rahardjo, 1993):

  $$\psi = -\frac{R\,T\,\rho_w}{M_w}\ln(RH) \qquad \text{[equation 80]}$$

  $$RH = \frac{p}{p_{o_1}} \qquad \text{[equation 81]}$$

  Where R is the gas constant (8.314 J/(mol K); T is the absolute temperature, M_w is the molecular mass of water (18.016 kg/kmol) and ρ_w is the density of the pure water (998 kg/m^3 at 293 K). A decrease in T should induce a decrease in ψ and, according to Cardoso et al (2007), oscillation of the RH due to the temperature dependence of p_{o_1}.

  Equation 80 and 81 can explain very well the observed RH oscillations due to temperature drop while the reason of the sharp increase in suction of sensors 1 and 5 does not stay in equation 80 and 81 but in the high dependence of the water potential readings to temperature below -500 kPa as reported by Teros 21 technical manual. Sensors 2, 3, 4 are in a condition of pwp above -500 kPa (low influence of temperature change on the readings) while sensors 1 and 5 below -500 kPa (higher influence of temperature).

- Over the whole period, all sensors experience some degree of fluctuation; mostly due to temperature fluctuations (see temporal period 125-135 days in Figure 88 and Figure 89).

3_7_2_3 Below the root bulb

In phase 1A and 1B, higher suction values are recorded by the sensors closer to the root bulb because of the increased water uptake. The last data of phase 1A show that the lowest value of suction is about 14kPa

(sensor 1) and the highest is around 250 kPa (sensor 8). Sensors 8 and 6 display close suction values and so do sensors 3 and 4.

Note that there is very little change in suction trend coming from dropping the relative humidity from ~99% to 74% at day 117. This is most likely due the fact that these sensors are directly under the root bulb and the potting mix and above the potting mix a cover of dry improved soil is located (150 mm of thickness). At day 117, sensor 10 (30 mm below the potting mix) is at 270 kPa of suction which corresponds to the inflection point in the bimodal SWRC of the potting mix (see Figure 61). As reported by Simunek et al (2012b) the hydraulic conductivity function has a slowdown in the permeability decrease in correspondence of the transition between microscopic and macroscopic porosity region. An example is presented in Figure 90. This effect on the permeability of the potting mix is covered by the very low value of permeability of the dry crust on the top surface that slowdowns the evaporation contribution. As a result the evaporation contribution at day 117 and in the days immediately following is very reduced.

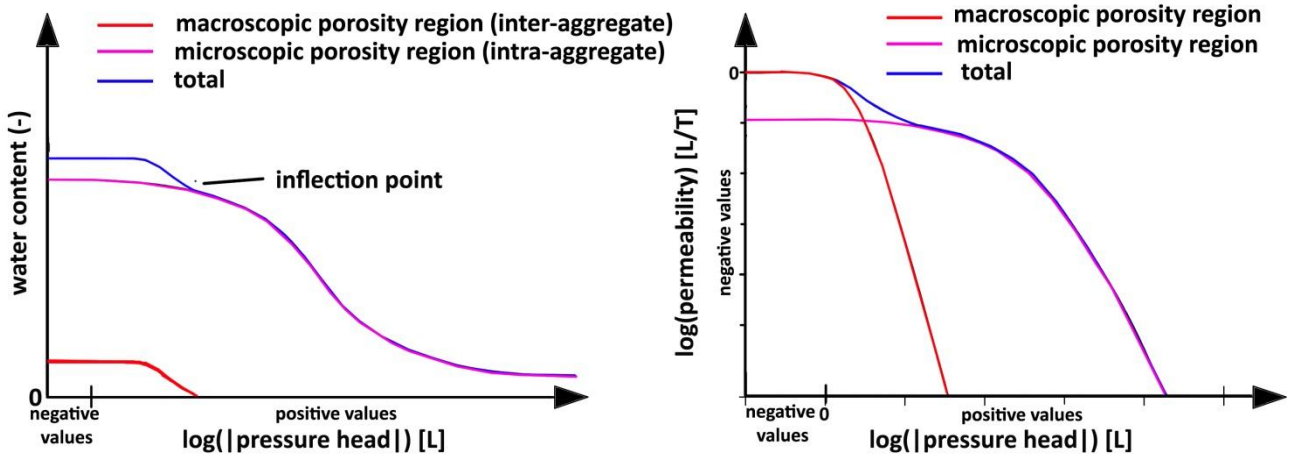


Figure 90: Example of composite SWRC (left) and hydraulic conductivity function (right) for a dual-porosity material as the potting mix.

If we compare the response of sensors 10 (below root bulb) and 13 (in the control zone, at same elevation), it can be seen that the suction at sensor is much higher than that of sensor 13. For example, at day 145 sensor 10 records ~300 kPa of suction against ~65kPa for sensor 13. Such significant difference is attributed to the root water uptake effect. In contrast, sensors 7 (under the root bulb) and 14 (in the control zone, at same elevation) return similar values of suction (e.g. 14kPa at day 80), which shows that, at that depth, suction is not affected by the root bulb.

Sensor 11 is at the same depth than sensor 9 but with a 145 mm radial offset (under sensor 4). Given that sensors 9, 2, 3 and 4 all record an influence of the root bulb, it is reasonable to consider that so does sensor 11. Its higher value is due to the fact that it is further away from the root bulb that sensor 9 and 4.

Small fluctuations are visible for all sensors, which is due to local variation of temperature (see Figure 88 and Figure 89).

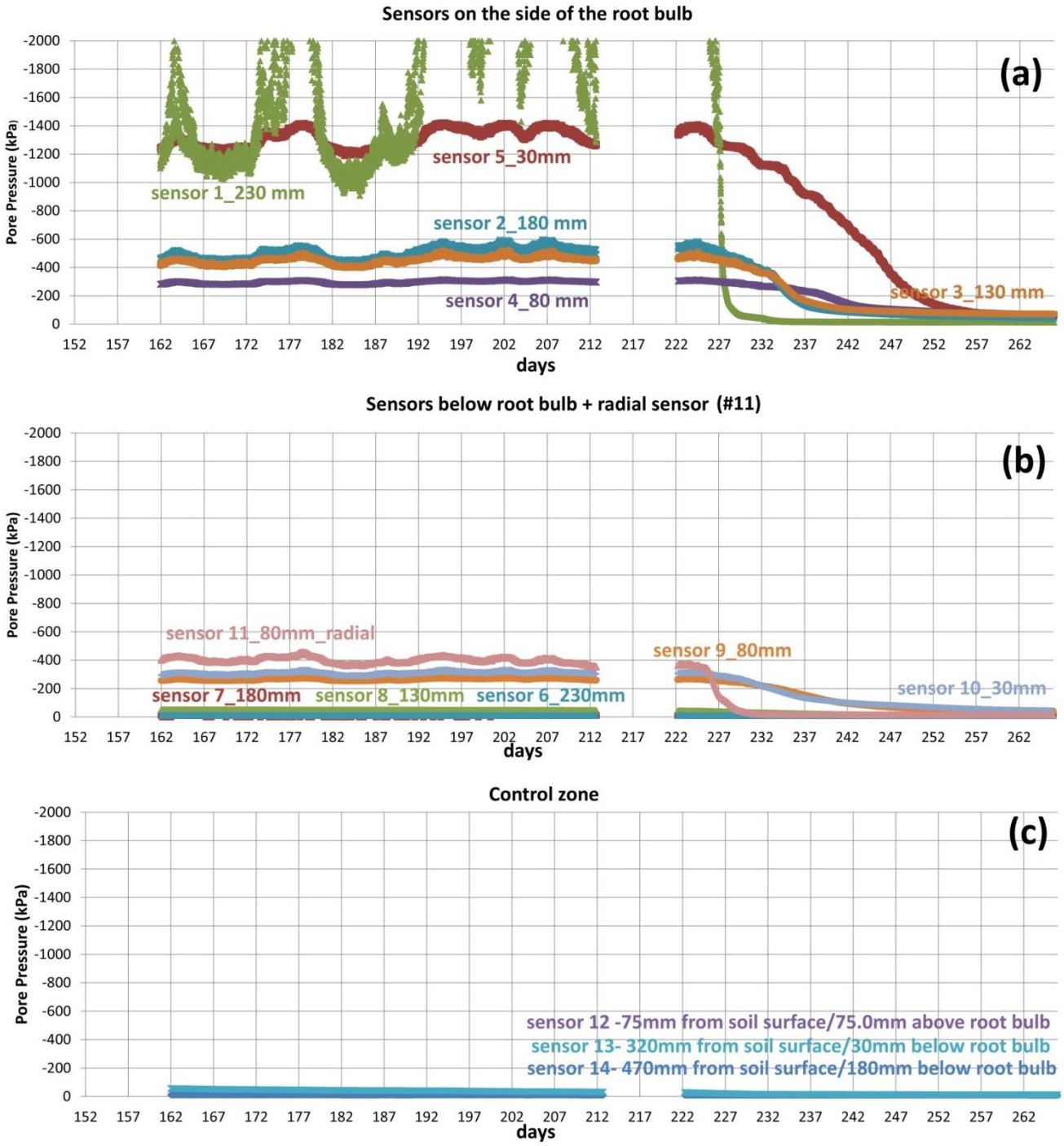


Figure 91: Evolution of pore pressure with time on the side of the root bulb (a), under the root bulb (b) and in the control zone (c) in phase 2 and 3. Water level is set to 495 mm from the soil surface in phase 2 and 418mm in phase 3. RH is set ~78% in both phases.

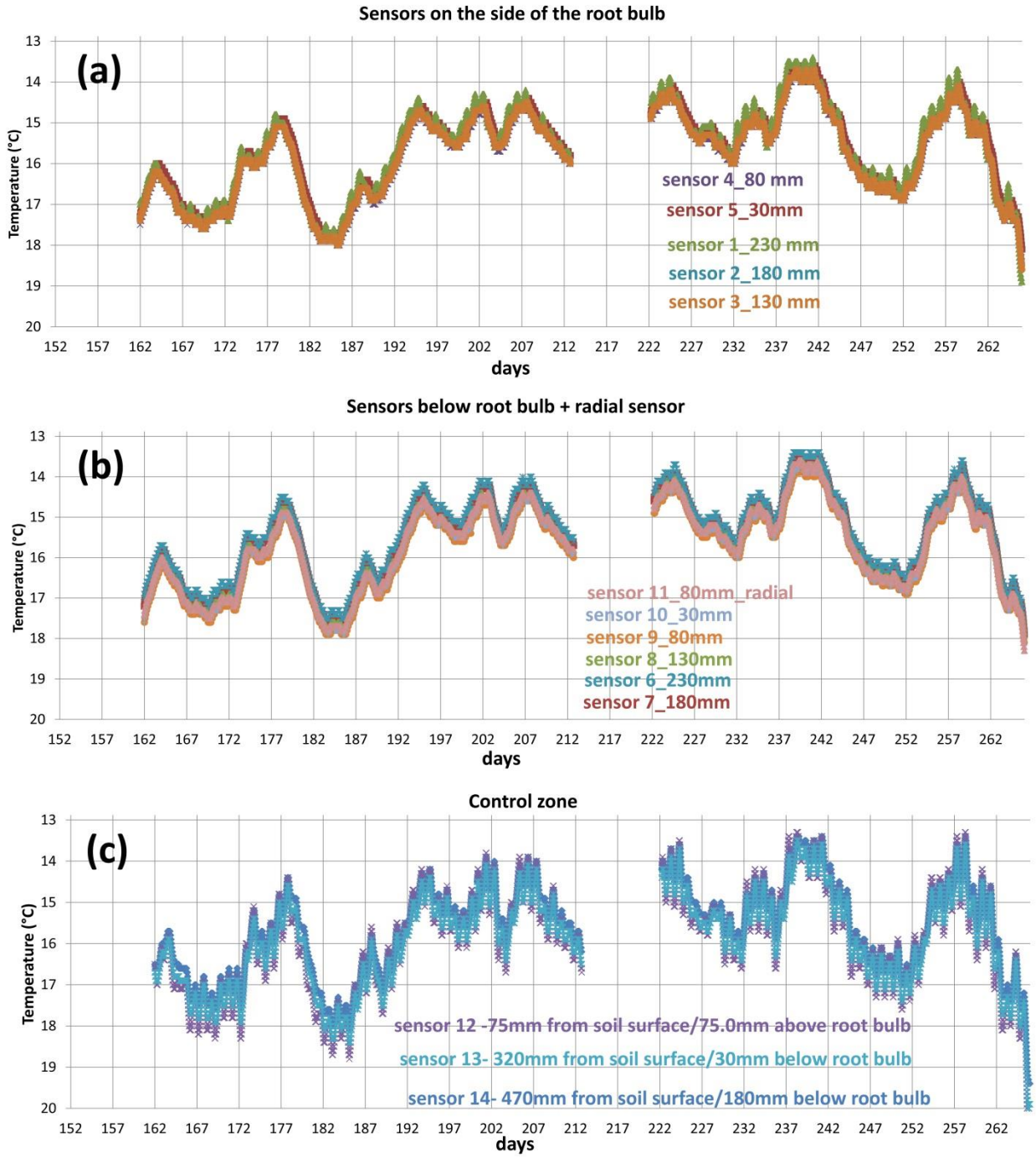


Figure 92: Evolution of temperature with time on the side of the root bulb (a), under the root bulb (b) and in the control zone (c) in phase 2 and 3. Water level is set to 495 mm from the soil surface in phase 2 and 418mm in phase 3. RH is set ~78% in both phases.

3_7_3 Phases 2 and 3

The data pertaining to phases 2 and 3 are shown in Figure 91 and the results are discussed zone by zone (control zone, on the side of the bulb, below the bulb). Please note that another logging issue occurred between days 145 and 162 and between days 213 and 222, so not suction or temperature was recorded over that period.

3_7_3_1 Control zone

At the beginning of phase 2 (day 152), the water level was raised to 495 mm from the soil surface (+155mm with respect to phase 1B). In the temporal range 162-213 days, suction of sensor 7 hardly changes and can be considered to be at equilibrium while suction of sensor 9 keeps reducing at about 0.50 kPa/day, suggesting that reaching equilibrium at that point is extremely slow. This trend continues in phase 3, where the water level has been raised further.

3_7_3_2 On the side of the root bulb

The order of suction readings remains the same between phase 1B and phase 2, where the main difference is an increase of water table position by 155 mm. Readings of sensor 1 are very affected by temperature oscillations showing peaks of suction over 2000 kPa but the overall trend (when ignoring the temperature fluctuations) is an increase in suction that started in phase 1B due to the evaporation contribution (RH=78%). There is a consistency between the temperature and pwp oscillations in time. In Figure 91 and Figure 92, the increase in temperature between163-170 days (from 16.0°C to 17.5°C) and 178-182 days (from 14.8°C to 18°C) of the sensors on the side of the bulb causes in both the cases a reduction of the suction values (particularly visible for sensors 1 and 5 due to their higher suction value above 500 kPa).

Sensor 5, 30 mm from the root bulb, shows a slow increase in the suction value (and temperature induced fluctuations), which is the result of the combined effect of evaporation and transpiration processes. Interestingly, at day 200 it is approaching the wilting point (1500 kPa), which means that the roots have not grown passed sensor 5, the roots will not be able to grow radially and extract water. Readings of sensors 2, 3 and 4 are largely unchanged from phase 1B.

In phase 3, the water table was raised by ~ 77mm. It takes about 12 days (220-232 days) after rising the water table to see the sensors report a significant drop of suction which lasts for about 20 days (232-252 days).

More specifically, sensor 5 drops from 1336 kPa to 65.7 kPa of suction in 44 days at an average rate of 29 kPa/day. In the same period (222-266 days) sensors 2, 3 and 4 see suction drop at rates of 11 kPa/day, 8.8 kPa/day and 5.7 kPa/day, respectively. The rates are computed as linear interpolation between initial and final suction values divided by the time period. Final values of suction are in the range 65to 6 kPa, for which the soil is close to saturation as observed in the SWRC of the improved soil in Figure 52.

Hypoxic or anaerobic conditions, in which the quantity of available oxygen is extremely low, causes a reduction or zeroing of the plant transpiration and nutrient absorption from the soil to the roots. A prolonged period of time in that condition could lead to roots death, great susceptibility to disease and at the end, eventually death of the whole plant. For this reason, phase 3 was terminated about 1 week after suction values drop close to 0kPa. The water table was subsequently dropped by 120 mm.

3_7_3_3 Below the root bulb

At day 152, the water level has been raised by 155mm with respect to phase 1B. Between days 162 and 200, all the sensors below the root bulb experienced a slight increase in suction as could be observed in Figure 91. The closer the sensor to the bulb, the higher the rate of suction increase. Suction of sensor 10 (30 mm below the bulb) increases at a rate of 0.65 kPa/day, that of sensor 9 (80 mm below the bulb) at 0.315 kPa/day; that of sensor 8 (130 mm below the bulb) at 0.078 kPa/day and that of sensor 6 (230 mm

below the bulb) at 0.02 kPa/day. Sensor 11 (at same depth than sensor 6 but with a radial offset of 145 mm) experiences a similar response than sensors 9 and 10, the closest to the bulb. However, it seems that the sensor not directly under the root bulb is more sensitive to temperature fluctuations. As already highlighted this is due to the sensitivity of the sensor to temperature which affects the capacitance reading at matric potential lower than -500 kPa. Among the sensors below the bulb, sensor 11 is at the lowest pwp (~-430 kPa). Sensors 7, 8 and 6 return the same readings than in phase 1B because they were already close to saturation.

In phase 3, because of the additional rise of water table, just like on the side of the bulb, sensors under the bulb return a rapid decrease of suction.

Comparing sensors 10 (30 mm below the root bulb) and 13 (control zone, at the same depth than sensor 10), it can be seen that not only suction of sensor 10 is higher than that of sensor 13 but the trends are different: suction slightly increases at sensor 10 while it slightly decreases at sensor 13, which is due to the transpiration effect of the plant.

As for sensors 7 (180 mm below the root bulb) and 14 (in the control zone, at the same depth than sensor 7), they both have reached equilibrium, at a suction of ~ 11 kPa.

3_7_4 Some considerations on the experimental data

Figure 93 shows the evolution of suction values with radial distance from the bulb (Figure 93b) and with depth under the root bulb (Figure 93a) at different times of the experiment.

Let us first analyse the evolution of suction on the side of the bulb (Figure 93b). Day 35 was taken as an "initial" condition, after three wetting events. The lower value of suction is recorded close to the root bulb due to wetting, while suction increases moving away from the bulb. After that, in time, the profile of suction changes due to two effects:

- evaporation through the soil surface shifts the curve on the right side
- root water uptake increases suction in the first sensor (located at 30mm from the bulb) to values very close to the wilting point, which is detrimental to plant growth.

The increase of suction at 80 mm from the bulb seems to be essentially due to evaporation, which suggests that the influence of the root bulb is only limited to 30 mm. That said, the actual size of the root bulb in the potting mix is unknown and has been assumed to be equal to the side of the pot in which the plant grew.

Looking at the suction profile under the root bulb (Figure 93a), the trend is monotonic:

- the closer the sensor to the bulb, the higher the suction, which is a clear effect of water uptake by the root bulb.
- the more time passes, the higher the suction under the bulb. Between 60 and 145 days, the sensor 30 mm under the bulb experiences an increase of 69.5 kPa while the one at 80 mm only sees an increase of suction of 40.8 kPa over the same period.

At that depth, the suction increase is mainly caused by the plant, not by evaporation because of the large amount of cover. The zone of influence of the root bulb seems to extend at least to 80mm, but not all the way to 130 mm. This possibly because roots extend preferably where water is (i.e. where suction is the

lowest) and water seems to be ponding under the potting mix. Indeed, values of suction are much lower under the bulb than on its side. The curve at day 35 in Figure 93a does not match the trend of the following ones due to the fact that in that day the third irrigation along the plant stem has been performed causing a reduction of the suction value close to the bulb (30 mm) and at day 30 the water table has been set to 650 mm from the soil surface causing a rapid decrease in suction of the deepest sensor (6).

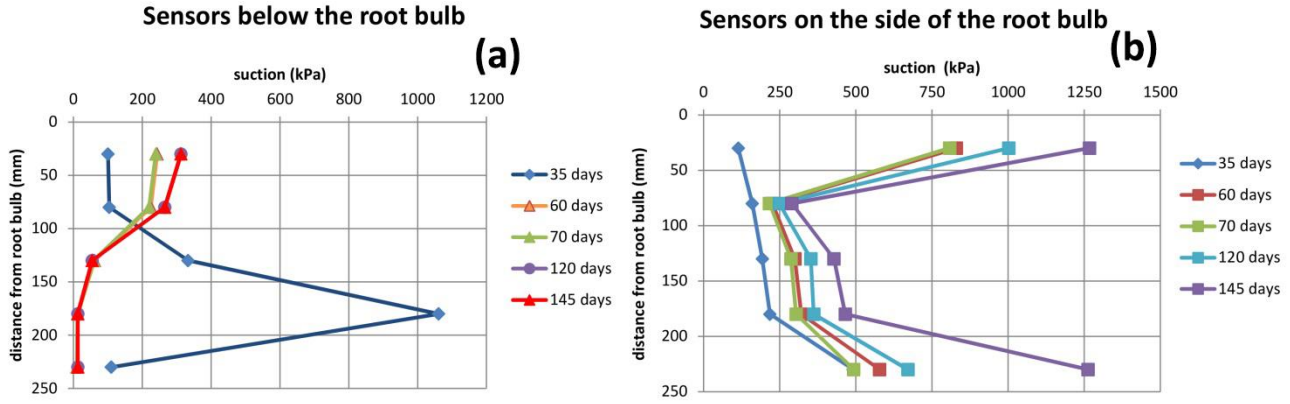


Figure 93: Evolution of suction values with distance from the root bulb at certain temporal instants of the test, for the sensors below the root bulb (graph a) and for sensors next to the root bulb (graph b).

3_7_5 Preliminary conclusions

The experimental program of the large scale apparatus is divided in 6 phases. In phase 0 the RH in the atmospheric boundary is uncontrolled, plant is watered three times along the stem, and the water table is applied to the bottom boundary just above the layer of gravel. In phase 1A the water table is raised with respect to the previous phase and RH set to ~99%, in phase 1B RH is lowered to ~74% while the water level is kept the same. In phase 2 and 3 the water table is raised again incrementally and RH is kept the same of phase 1B. Phase 4 sees the last rising of the water level and the corresponding flood of the root bulb (RH ~99%). In phase 0 the sensors on the side and below the root bulb oscillate around their initial pwp value in equilibrium with the surrounding soil. Oscillations decrease moving away from the soil surface in fact the soil provides a cover against temperature variations. It has been observed that if the pwp readings of the sensor (in negative sign) are below 500 kPa, the temperature effects on the measurement readings increase. In this phase the soil around the root bulb is close to saturation (due to the irrigation along the stem) and the plant right after the re-planting starts to absorb water. Sensor 5 (30 mm from the side of the plant) shows an increment of 600 kPa in two weeks while sensor 10 (30 mm below the root bulb) an increment of 200 kPa. This could be due to a different root density between the side and below the root bulb or a ponding water effect on the bottom of the root bulb at the interface between improved soil and potting mix. In the same period of time the remaining sensors on the side and below the root bulb show an increase in negative pwp much lower (~100 kPa) due to evaporative effects from soil surface. In phases 1A and 1B analogous observations could be done : the RWU effects are observed only for sensor 5 (30 mm on the side of the bulb) which is affected also by evaporation from soil surface, while the remaining sensors on the side of the bulb show an increase in suction due only to the evaporative phenomenon. These information could be extrapolated comparing the response of the sensors close to the plant and the sensors at the same depth but in the control zone (no influence of the RWU). If we consider the sensors below the root bulb, the RWU phenomenon is observed only for sensors 10 and 11 (30 and 80 mm from the soil surface, respectively) and, due to the thickness of the soil cover above them, evaporative effects can be excluded. The greater depth of influence of the RWU below the root bulb could be due to a growth of the

roots toward the water table (downward). We can conclude that the plants effects on pwp distribution in soil could be observed only in a restricted soil volume close to the root bulb (30-80 mm from the initial dimension of the root bulb assumed with the shape of the pot in which the plant grew in the nursery). A long equilibration time is observed for each experimental phase in which boundary conditions have been changed. The water level has been raised each time step in order to avoid an excessive depletion of water in the soil volume close to the root bulb and increase the available water for the transpirative phenomenon. In phase 1B the decrease in the RH of the upper boundary from ~99 to ~74 is not accompanied by an increase in the evaporation rate due to the creation in phases 0 and 1A of a dry crust of soil which caused a drastic decrease in soil permeability according to the HCF. We can conclude that, after phases 0 and 1A, the evaporative phenomenon has a reduced (or null) relevance on the experimental results.

4 CONCLUSIONS PART II

The PART II of the present thesis investigates the evapo-transpiration effects on the spatial and temporal distribution of pwp in close proximity of the root bulb of a Malaleuca Styphelioides, native Australian evergreen plant. The study moves from stability problems of an embankment located behind the retaining wall that borders the Newcastle Inner City Bypass Cut 7, West Charlestown (Newcastle, Australia). The use of long-stem planting has been suggested to solve the shallow instabilities of the slope (for more information see chapter §1). The physical characterization of the in situ soil has been performed in order to individuate the best planting solutions. It has been observed that the in situ soil is an organic clay of high plasticity, low salinity (as desirable) but extremely acidic (pH ranging from 4.8 to 5.4) which is detrimental for plant growth. Moreover the saturated hydraulic conductivity of the soil is extremely low ($5.6*10^{-8}$ m/sec) which means a soil behaviour dominated by micro-pores and a lower water availability for plants.

To improve the conditions for plant growth, clean river sand is added to the original soil in order to increase the saturated permeability ($2.93 \cdot 10^{-7}$ m/s) and decrease the air entry value. The new composition of the soil is 65% coarse material, 35% fine material. Lime is added to the original soil to increase the pH of the soil to 6, optimal soil acidity. Improved soil is going to be substituted only in the points of the slope where high-stem planting will be performed (for more information see chapter §3_1_1). The soil water retention parameters (SWRC) of the improved soil have been investigated in terms of gravimetric water content (by means of tensiometers readings above the cavitation pressure and dew-point potentiometer readings below that point) and volumetric water content (by means of hand spray plaster method developed by Liu and Buzzi (2014)) versus water potential (see chapter §3_1_2_2). The microstructure of the improved soil has been investigated by means of mercury intrusion porosimetry (MIP) which revealed that the retention behaviour is dominated by macropores. The saturated permeability of the improved soil has been investigated by means of a Rowe cell (chapter §3_1_2_3).

Laboratory tests have been performed also on the potting mix, highly organic non-cohesive material in which the plant has been grown in the nursery (see chapter §3_1_3). The investigation of the SWRC (suction vs gravimetric water content) revealed that potting mix is a dual-porosity material. The SWRC in terms of volumetric water content has not been investigated due to the non-cohesive nature of the soil that makes impossible to monitor volumetric deformations upon drying. The saturated permeability of the potting mix has been investigated by means of a constant head permeability test.

A 1D infiltration experiment has been conducted in order to investigate the parameters of the main wetting curve and of the hydraulic conductivity function: a column of material (improved soil) close to dry conditions is subjected to an upward gradient of water which is fed to the system with a constant volume from the bottom boundary by means of a GDS controller. Suction is monitored at two different heights in the column by means of high capacity tensiometers. A sensitivity analysis of the hydraulic parameters has been performed in order to exclude from the subsequent calibration phase the parameters to which variations the model is insensible. The θr parameter has been excluded for a sensitivity of the numerical model to parameter change close to zero. It has been observed that the sensitivity of the parameters is limited to the temporal period from the arrival of the hydraulic stimulus in the considered observation point to the complete saturation, while in the remaining period is zero. A calibration programme in two phases, simplified with respect to the one used in PART I of the present thesis, has been carried out. 30 inverse analysis have been performed starting from a set of hydraulic parameters from laboratory test and starting from an initial guess of the parameters using pedo-transfer functions. The suction trends with time in the two observation points have been used in the indirect problem for half of the inverse analysis and for the other half suction measurements and the initial water content of the soil. As done in part I, hydraulic parameters have been optimized singularly, in pair, in triplet up to all the parameters together. As observed also in the calibration performed in part I, the best performance (best match observed-simulated datasets) has been obtained using pressure head observed data and the initial water content and optimizing a low number of parameters (in this case θs ,α and n). A possible evolution of the equipment is the installation of a sensor monitoring water content in close proximity of a tensiometer in order to trace the whole local SWRC.

The project focuses on the design and realization of a large scale apparatus which consists of a 1 m^3 of improved soil with a Malaleuca Styphelioides plant in the centre. 6 sensors monitoring pore water pressure (Teros 21 by Meter group) have been installed below the root bulb, 5 on the side of the bulb and 3 in a control zone in one of the edges of the apparatus separated with an impervious material from the plant. The potting mix in which the plant has been grown could not be separated from the root bulb (intricate maze of fine roots and soil particles) so it is inserted as it is in the centre of the apparatus. A water level is applied to the bottom boundary of the apparatus and RH is controlled in the atmospheric boundary. Different water levels and RH combinations have been tested.

6 different stages have been applied to the large scale apparatus: stages 0 to 3 see the application of an increasingly high water level to the bottom boundary and stage 4 see a reduction in the water level with respect to stage 3. A variable RH is applied to the upper boundary and it is equal to 99% for phase 1A and 4, and ranging from 70-80% for phases 1B, 2 and 3. For more information see chapter §3_6. In phase 0 three irrigation events were performed along the plant stem in order to provide the necessary water for the transpiration demand before the application of the water level.

The experiment has been reproduced numerically using the commercial code Hydrus 2D able to simulate water movement in variably saturated media taking into consideration the uptake by plant roots (chapter §2_1_2). The van Genuchten model has been adopted to represent both the retention behaviour of the investigated materials (potting mix and improved soil). Although this choice is not the best solution for a dual-porosity material as the potting mix, it has been preferred to keep a lower number of hydraulic parameters in order to ease the calibration of the model. The parameters to be calibrated are the hydraulic parameters of the SWRC and the hydraulic conductivity function which is extrapolated according to Mualem's model (6 parameters); the parameters of the water stress response function of Feddes et al

(1978) (7 parameters); the parameters of the root water uptake spatial distribution function of Vrugt et al (2001) (6 parameters).

The sensitivity analysis applied to the model shows that the RWU spatial parameters are the ones to whom the model is more sensitive, while Feddes' parameters and SWRC parameters have a much reduced relevance for the model output. Moreover, as observed for the 1D filtration experiment, the sensitivity of the model to parameter variation is observed solely during the simulation period between the arrival of the hydraulic stimulus and the complete saturation of the observation point.

The marginal influence of the retention parameters is confirmed also after the calibration carried out by means of inverse analysis. The same simplified procedure adopted for the 1D filtration experiment in column is proposed again. The performed calibration optimises the performance of 2 out of 14 observation points and the variations in the hydraulic parameters are observed only for α, Ks of the potting mix. When more than one set of hydraulic parameters has to be optimized together, the inverse analysis application shows limits, as observed also in in the calibration program performed in PART I of the present thesis. The enhancements of the model performance are often localized in one single material and rarely groups of materials optimized together give the greatest enhancement among the performed indirect simulations. This is due to the fact that problems of uncertainty, non-identifiability and correlation of the optimized parameters increase increasing the number of parameters to calibrate.

Hydrus allows taking into consideration the compensatory RWU mechanism which considers that the plant may respond non-uniformly over the root zone to water stress. The reduced amount of water uptake in the most water stressed part of the root zone is compensated by an increase of uptake from the less stressed parts. Hydrus does not allow calibrating the water stress index ω_c using the inverse analysis, so a trial and error procedure has been used to individuate the water stress index which reduces the differences between observed-simulated datasets in the observation points. A qualitative evaluation by means of graphics of pwp with time and a quantitative evaluation by means of a set of metrics/indices have been used to analyse the different model outputs. A critical water stress index of 1 results from the performed calibration that means no compensatory ability of the investigated plant.

Hydrus does not allow calibrating the RWU spatial parameters of Vrugt et al (2001) by means of inverse analysis, for this reason a trial and error procedure has been performed in order to obtain the best match between simulated and observed dataset in the observation points. The RWU spatial distribution of Vrugt et al (2001) has been compared to a spatial distribution obtained without the use of any predetermined function but tuning the local dimensionless root distribution parameter $b(x,z)$ in the observation points using the pwp in situ information and interpolating in the remaining nodes of the rooting domain. The description of the RWU spatial distribution by means of the Vrugt et al (2001) function leads to many constraints in the geometry of the RWU pattern and a poor prediction of the local uptake in close proximity of the root bulb. The elaboration of the RWU spatial distribution from the interpolation of the dimensionless $b(x,z)$ parameter in the observation points was found to be a better alternative.

Roots extract water at higher rate from the upper part of the soil profile then gradually less moving to deeper depth, in accordance with Deb et al (2013). This is due to a different distribution of the root density with depth. This is observable also in the present case study where sensor 5 (30 mm on the side of the root bulb) extracts water at higher rate than sensor 10 (30 mm below the root bulb). The same conclusion could be drawn looking at the calibrated local RWU spatial distribution parameter $b(x,z)$ equal to 0.019 below

the root bulb and 0.051 on the side of the root bulb. As expected, the extraction of water by roots decreases moving away from the centre of the bulb and $b(x,z)$ reaches very low values from 180 mm on the side of the root bulb in the horizontal direction and from 130 mm in the vertical direction.

Looking at the simulation of the calibrated model, as expected, water extraction by roots first affects the parts of the domain closer to the bulb, which are depleted of water. Then, a drying front propagates outwards due to a hydraulic gradient from the farer wetter regions to the closer (to the root bulb) drier region and, in the meantime, a plant would grow fine roots to reach water and nutrients.

Looking at the five observation points on the side of the root bulb, they are subjected equally to the evaporation contribution and differently to the transpiration contribution. RWU effects are observable clearly for sensor 5 positioned at 30 mm from the side of the bulb while the increase in suction of the sensor at 80 mm is due only to evaporation. RWU in observation point 5 (30 mm on the side of the bulb) is extremely relevant, after the last irrigation along the plant stem in 15 days suction has increased of 600 kPa which is half the wilting point and in 180 days it has reached -1400 kPa (wilting point). Looking at the six observation points below the root bulb, it has been observed, as expected, that closer the observation point to the root bulb higher the suction values recorded due to an evident plant influence. The influence of the root uptake is observable up to 80 mm from the root bulb. In the initial phase, irrigation along the stem guaranteed an high water content in proximity of the bulb and an higher uptake on the horizontal direction with respect to the vertical direction due to a probable ponding of the water at the interface between the very permeable potting mix and the less permeable improved soil which causes a lower availability of water for the sensors below the root bulb. With time, as suction increases notably on the volume around the root bulb, roots start protruding toward the water table and the soil volume below the root bulb already interested by the upward hydraulic gradient. Observation points 10 and 9 (30 and 80 mm below the bulb, respectively) maintain lower suction value due to a higher availability of water with respect to the observation points on the side of the bulb (positioned farer from the imposed water table).

Temperature variations have a relevant influence on the suction trend of the sensors on the side of the bulb when suction values are greater than 500 kPa while for values below this threshold the influence is low or null. In order to eliminate any temperature oscillation from sensor readings a possible future development of the large-scale equipment could be the regulation of the temperature at the upper boundary. The sensors below the root bulb, due to the high soil cover do not show to be influenced significantly by temperature variations.

The evaporation contribution in the large-scale experiment is only marginal. During phase 1A, the RH is ~100% and the evaporation is close to zero. A peak of evaporation has been observed in phase 1B because the RH has been lowered to 75%. In time, due to the desiccation of the external surface which led to the creation of a crust and a rapid decrease in the hydraulic conductivity of the soil, the outflow dropped from values around 0.00020 m/day to 1-2*10^{-6} m/day. After that, the outflow due to evaporation is pretty constant in time because the surface is inside an enclosed space, out of direct sunlight so there is no cyclic behaviour due to alternation of day and night.

A possible future development of the present work is the replication of the same experiment using different plant species in order to individuate the ones that are more appropriate for the in situ planting and allow an easy comparison between them under the same conditions. The final purpose is the elaboration of a reliable RWU spatial distribution to be implemented in a more complex numerical model of

the investigated slope able to take into appropriate consideration the evapo-transpiration phenomenon which has relevant influence on slope stability.

GENERAL CONCLUSIONS (PART I, PART II)

Chapter 10 of PART I reports the main conclusions regarding the process of optimization of the model parameters of the instrumented embankment section along Secchia River (period 2018-2019) and the quantitative evaluation of the indirect simulations performance using a methodological approach in three phases. In chapter 4 of PART II of the present thesis, the main results obtained from the experimental development of a large-scale apparatus and the numerical modelling of the laboratory experiment are presented extensively. The present chapter wants to emphasize the degree of similarities and differences between the two parts of the thesis and the main general conclusions which may be drown at the end of the discussion of both the topics. Together with these aspects, a more critical overview on the obtained results is presented highlighting potentialities and limits, pros and cons of both the presented topics.

Main differences between the goals of PART I and PART II

As underlined many times, the two topics have different final goals although intrinsically related. Both the topics deal with unsaturated soil mechanics and transient seepage phenomena but in one case (PART I), the scale is the in situ one (large scale), while in the other case (PART II) the scale is the one of the laboratory (small-medium scale). This corresponds to a different degree of representativeness of the investigated phenomena and accuracy of the model simulations.

PART I focuses on the optimization algorithm of Marquardt-Levenberg and its application to the calibration of transient seepage models while in PART II model calibration is just a means to obtain a numerical model able to investigate more deeply the relation between evapo-transpiration contribution and suction distribution in close proximity of a plant root. In fact inverse analysis technique has been used in PART II to calibrate the hydraulic parameters while RWU parameters, water stress function parameters and parameters of the compensatory mechanism have been calibrated using a ‘‘trial and error’’ procedure.

In PART I, the calibration procedure applied to the numerical model is complex and time consuming due to the fact that the number of hydraulic parameters to be optimized is high not only because we are dealing with a multi-layered domain but also because different casuistries of hydraulic models, weighting distributions, observed datasets have been tested in order to investigate in detail the calibration phase itself in order to elaborate using this wide database of indirect simulation a methodological approach for the evaluation of their performance. The procedure of calibration applied to the hydraulic parameters of the model in PART II is much simpler because the model is less sensitive to the variation of the SWRC parameters and a lower number of parameters has to be calibrated (only two layers, one hydraulic model). In this case, the methodological approach in three phases used to evaluate the performance of the indirect simulations in PART I, has been simplified in two phases and a lower number of indices/metrics has been taken into consideration.

In both the topics, the evapotranspiration contribution is considered but with a different degree of accuracy: in PART I, due to a lack of input data, simplifications have been adopted, while in PART II, all the experimental data were available and the evapotranspiration phenomenon is dealt with higher accuracy.

While in the numerical model of PART I, the vegetative cover of the river embankment has been dealt as a transpiration contribution distributed over the atmospheric boundary, in the numerical model of PART II the root water uptake distribution depending on the roots spatial distribution has been specified in order to consider the single plant and the local effects of the transpiration. From all this, we can understand the different scales of the problem: from the vegetation cover considered as a whole (riparian bushes of PART I) to the single plant (medium-long stem plants of PART II).

Potentialities PART I

The evaluation methodology elaborated in the calibration procedure is easily and directly applicable to earth structures subjected to transient boundary conditions as dams and embankments used for road and rail infrastructures. It is relevant to highlight that changes have to be performed on the chosen weighting distributions adopted. The modeller needs to fix prior to the calibration phase which are the intervals of observation data more relevant for the purposes of his analysis and choose appropriately the weighting distribution. For example for our purposes the medium water levels more persistent in time are the hydraulic events most feared for riverbank stability, for this reason the tested weighting distributions give more weight to these temporal periods. In the same way, in the phase of evaluation of the indirect simulations performance, points given to the different metrics/indices have to be subdivided proportionally in the different simulation periods based on the same logic.

It has to be underlined that one part of the metrics/indices/properties used in the calibration procedure exposed in PART I can be applied to evaluate any kind of numerical model which disposes of an observation dataset with which to evaluate the discrepancies between observed-simulated predictions. The remaining part of the metrics/indices/properties can be applied only to the hydraulic-geotechnical modelling.

The great availability of data (three years of acquisition) is the major potentiality of the research because it allows to use one year (2016-2017) of data for the calibration of the initial conditions, one year (2017-2018) to run the direct simulation and calibrate the hydraulic parameters by means of inverse analysis and one year (2018-2019) for the validation of the calibrated model. Moreover 25 sensors monitoring water content and pressure head in the same riverbank section provide complete information on the pwp changes in space and time.

Potentialities PART II

The experiment conducted in the large scale apparatus on the Malaleuca Styphelioides plant can be replicated easily with other plant species in order to investigate different types of root bulb architectures and different RWU spatial distributions. This allows individuating the best solution for the investigated soil of the instable embankment slope.

The large-scale apparatus could be easily replicated in other laboratories because it uses commercial sensors which need no user calibration, they are relatively cheap and easy to install. This allows to compare results obtained with different plant species and to create a database.

The experiment has relevance in all cases in which a numerical model needs an accurate computation of the transpiration contribution of plants because it guarantees all the necessary information to calibrate the transpiration parameters. Moreover the procedure of calibration of the RWU spatial distribution

parameters using the information in the observation points (where sensors have been installed) is simple and easily applied.

The use of installed pwp sensors (as Teros 21 by Meter Group used in the large-scale apparatus) for the calibration of the RWU spatial distribution could be a valid alternative to destructive techniques which involve the sampling of cylinder of soil around the root bulb in order to obtain information with depth on the root length density with which to elaborate RWU spatial distribution. Moreover due to the fact that the root architecture changes in time for natural growth and external stimuli, the RWU spatial distribution needs to be updated periodically performing each time other root samplings which affect the root structure. The use of installed instrumentations to measure pwp in relevant points allows collecting data in order to update roots architecture in time with no additional costs. The same sensors and the same disposition of the instruments could be applied in situ to investigate adult plants of considerable dimension (root bulb diameter and plant height). The choice of the field scale if in one hand allows having the best representation of the in situ conditions of the living plant, in the other does not allow to control boundary conditions as the relative humidity at the upper boundary. Set a RH close to 100% means to separate the evaporation contribution from the transpiration contribution and this simplifies largely the interpretation of the experiment results.

Limits PART I

For what concerns the elaborated model of the river embankment section, it has to be underlined the lack of proper observed data in order to compute with accuracy the evapo-transpiration contribution from the atmospheric boundary. Moreover the atmospheric data are collected from a meteorological station 7 km away from the instrumented section and this introduces an error in the simulation (especially in the rainfall data which are affected greatly by the location). Moreover to compute with accuracy the changes in the water table height on the right side of the embankment section (far field) a piezometer would have been useful. In the elaborated model the water table height of the far field has been hypothesised constant.

Initial conditions proved to have a relevant influence on the model simulation. Differences between observed/simulated datasets in the initial instant of the simulation (initial conditions) are difficult to reduce or cancel in time (high water events are able to cancel partly these differences). The use of punctual data of PH and wc in the observation points (where sensors have been installed) and an interpolation in the remaining nodes of the model could be a valid alternative to reduce the gaps between observations and simulations as suggested by Gottardi and Gragnano (2016) and Gragnano et al (2019).

While a good convergence of indirect simulations has been observed in PART I starting from a dataset of hydraulic parameters from laboratory tests, great problems have been detected using different initial sets of parameters. These problems of convergence are result of model complexity and for this reason they have not been observed in PART II which deal with a simpler model (lower number of parameters to optimize, fully controlled boundary conditions, shorter simulation period).

It is thus important to stress that the enhancement in the model performance starting from the set of laboratory data is quite limited due to the fact that the performance of the model using the initial set of parameters is good and for this reason small space is left to its performance optimization. Subsequently the reduced improvements do not justify the calibration programme performed with its great number of indirect simulations. Despite this, it has interest for research purposes because it considers different

casuistries (different hydraulic models, different initial sets of parameters, different typologies of observed datasets etc).

The elaborated methodological approach in three phases used to evaluate quantitatively the performance of indirect simulations is quite time-consuming and it requires a high number of performed simulations. It is possible to skip the pre-calibration phase and simplify the first calibration phase in case of a simple model and a small number of parameters to optimize. In these cases, the attention could be concentrated on the observed-simulated comparison in the most interesting observation points of the bank section (III phase) and on the result of the calibration phase and the optimized set of parameters (II phase) while in the first phase a lower number of metrics/indices could be used and a lower number of simulated time period considered (peak/off-peak periods).

Limits PART II

One of the main limits of the large-scale experiment is that the equilibration time after each raising or lowering of the imposed water table is long due to low soil permeability. This causes a considerable increase in the experiment duration at each step.

From the elaboration of the sensors readings the influence of the temperature is quite clear in the shallow sensors (30 cm below the soil surface). In the large scale apparatus the RH was controlled and recorded at the upper boundary while the temperature was recorded but not controlled. Variations in temperature between day and night are quite relevant in the laboratory where the large-scale apparatus is stocked. An improvement of the setup is the control of the temperature at least at the upper boundary while the greater soil cover on the side of the box is enough to avoid the influence of the temperature variation on the installed sensors.

The performed experiment has involved the investigation of the RWU spatial distribution of a young Malaleuca Styphelioides (2 years old). Due to natural growth of the plant and of its root bulb, the experiment needs to be replicated at different plant growth stages. The chosen plant could rich up to 20m of height in the adult age; this makes impossible the testing at the laboratory scale and it requires moving the testing on a field scale.

General conclusions PART I

It has been observed that the application of the inverse analysis procedure to a complex multi-layered model simulating a medium-long period (as the one in PART I) is demanding in terms of number of indirect simulations to perform. Moreover the increase in the quality of the performance of the inverse analysis is in some cases limited, partial or insufficient if compared to the time required to run the whole pool of indirect simulations.

The major problems stay in the difficulties in the convergence of such indirect simulations caused by problems of non-uniqueness of the solution and ill-posed inverse problem. A possible solution that could partly solve such limits is to focus the analysis on a limited number of layers, the most significant ones for the simulation and on the more sensitive hydraulic parameters (as per indication of the sensitivity analysis) and use in input observation points able to be relevant for the inverse analysis (in meaningful simulation periods).

The L-M algorithm is more suitable for simpler models (as the one in PART II) and for which a lower number of parameters has to be computed, shorter simulation periods, controlled boundary condition and lower number of observation points to monitor. A further evidence of this is that the 1D filtration in column (PART II) for its simplicity (one single layer) has proven to be well and easily calibrated by means of inverse analysis.

It has to be stressed that both for complex or simple models, the use of inverse analysis by means of Hydrus2D to perform the calibration procedure, is very time consuming because it requires that the modeller sets each indirect simulation indicating the set of initial parameters, the parameters to optimize, their intervals of allowed variation and the observed dataset specifying a weight for each information.

For complex models other optimization algorithms are suggested as genetic algorithms (GA) which are becoming increasingly studied and used for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. These optimization algorithms are not yet implemented in commercial codes able to run both the direct and indirect simulations of water movement in saturated/unsaturated material but they have to be run externally from the program by means of programming language as Phyton and Matlab. This complicates the application of the GA by a wide audience of possible public and private users.

It has been observed that the use of an initial set of parameters from laboratory tests guarantees a lower calculation time, a more efficient inverse analysis application and a higher convergence percentage of the performed indirect simulations with respect to the use of different initial sets (for example applying pedo-transfer functions or using hydraulic parameters from literature).

It is really important that the modeller tunes a direct simulation of the observed phenomena (using the initial set of parameters) whose calculation is quick and smooth, prior to the application of the inverse analysis in the calibration phase. To do so, the right mesh dimension (and mesh refinements), the right time step increment, the right tolerance of the solutions; have to be set by the modeller in order to ease the convergence. This guarantees fewer problems in the obtaining of the optimized parameters in the indirect approach.

General conclusions PART II

PART II investigates the evapo-transpiration effects on the soil volume around a root bulb of Malaleuca Styphelioides. From pwp monitoring, it has been observed that the RWU is not homogeneous around the bulb but greater on the side of the plant with respect to the vertical direction. This could be due to a ponding of the water at the interface between potting mix and improved soil and to an elongation of the finer roots in the horizontal direction probably during the initial phase of irrigation from the top (along the plant stem). Initially the depletion of the water is concentrated on the soil volume immediately close to the root bulb, then a gradient is established from the wetter regions to the dryer region close to the plant. The sensors monitoring pwp show the same trend: the highest suction value for the sensor closer to the bulb and gradually lower suction values moving away from the bulb.

In some testing phases, evaporation and transpiration phenomena have been considered together, in other phases separately (evaporation brought to zero, setting the RH equal to 100% at the atmospheric boundary). Evaporation effects have been observed only in the shallower sensors (30 cm from the atm boundary) while for the sensors below the root bulb the effects are negligible. After the creation of a dry

crust on the upper boundary, the evaporation effects from the soil surface were negligible because a very low permeability was reached (according to the HCF).

The numerical model requires the calibration of a great number of parameters: the hydraulic parameters, the parameters of the stress reduction function and the parameters of the RWU spatial distribution function. The hydraulic parameters of the SWRC have been optimized by means of inverse analysis and it has been observed that their influence on simulation results is very limited in agreement with the results of the sensitivity analysis.

Despite this the calibration of the hydraulic parameters by means of inverse analysis results simpler and with shorter calculation times with respect to the one performed on the numerical model of PART I. In fact a simpler numerical model with a lower number of layers and parameters to optimize and a shorter simulation period, assures a higher percentage of successful simulations and a set of optimized parameters with better performances (lower residuals).

As observed in PART I, great benefits are identified using as initial set of parameters, the one obtained by laboratory tests in order to reduce the efforts in the calibration phase. Hydraulic parameters have been obtained in laboratory applying the experimental infiltration column and high capacity tensiometers to track the main drying branch of the SWRC and the Hand Spry Method (Liu and Buzzi, 2014) to investigate the volumetric changes with water content of the improved soil.

While hydraulic parameters show reduced effects on model output, the RWU spatial distribution parameters have a great influence on the simulation output but the use of spatial function as the one of Vrugt et al (2001) shows some limits in the representation of the real phenomenon. For this reason, instead of using parametrized spatial functions from literature, the RWU spatial distribution has been determined using the pwp observed data. The local dimensionless root distribution parameter b(x,z) is set by the modeller in the observation nodes in order to reduce the discrepancy between observed and simulated data and in the remaining nodes by applying an interpolation between the calibrated values of b(x,z). This methodology assures the best match between simulated/observed data.

Future developments of PART I

A possible future development of the presented work in PART I could be the elaboration of a calibration programme using Genetic Algorithm (GA) in order to individuate pros-cons of this optimization algorithm and compare with the results of the calibration performed using L-M algorithm. The same methodological approach in three phases for the quantification of the indirect simulations performance could be applied to the new pool of indirect simulations obtained using GA in order to have a final quantitative evaluation of the optimization reached.

Due to the fact that the acquisition in the instrumented section is proceeding, another year of observation data is now available and could be used to evaluate the performance of the calibrated model on a new monitoring period (2020-2021). The focus could be on the performance of the set of optimized parameters as the temporal distance between the time period used for the calibration and the new simulation period increases. It is interesting to investigate if the performance of the simulation worsens as this distance increases. The exposed methodological approach in three phases to evaluate the simulation performance could be applied also to the new simulation period. Moreover it is relevant to investigate if changes in the performance are detected between hysteretic and non-hysteretic models and especially if hysteretic

models are able to contain this drift in the model performance as time proceeds. This makes us understand if a model calibration could be done one single time or it needs to be updated from time to time. This becomes increasingly interesting for a long-time monitoring of river banks and a long-time simulation of the most critical sections. Another interesting aspect to investigate is the pros and cons of a long-time simulation period (f.e. from 2017 to 2021) instead of a yearly simulation with a recalculation each time of the initial conditions using for example the method suggested by Gottardi and Gragnano (2016). What is interesting to notice is the influence of updated initial conditions on the performance of the following simulation period. In other words, it is relevant to investigate if long-time simulations (years) need periodically a recalibration of the model parameters (hydraulic parameters) and/or a reset of the initial conditions to zero the discrepancies between observed data (wc/PH) in the observation points and simulated data.

Future developments of PART II

A possible future development of the project in PART II is to test the large-scale apparatus with different plant species in order to compare the obtained distribution of pwp in time and select the most appropriate species for the in situ planting.

Another possible future development of the large-scale experiment is the testing of plant species without the imposition of different water levels but using controlled irrigation from the top (ground level) by means of an irrigation system (as a sprinkler system). This allows to test the RWU around the cultivated plant and to observe the different developments of the root bulb architecture due to different external stimuli. The imposition of a water level below the root bulb forces the plant to elongate the roots downward while irrigation from the top forces a roots elongation in the horizontal direction.

It is also relevant to investigate the development of the underground architecture of the investigated plants as growth proceeds because, as can be well understood, plant age has great influence on transpiration contribution and on RWU spatial distribution.

Another interesting development is the incorporation of the elaborated RWU spatial distribution close to the investigated plant in the numerical model of the Newcastle Inner City Bypass, West Charlestown in order to compute carefully the pwp distribution in proximity of the long-stem plantings. The elaborated pwp distribution could be imported as initial conditions to perform reliable stability analysis able to consider appropriately the plant transpiration benefits on the stability conditions.

The large scale apparatus could be integrated with sensors monitoring water content in close proximity of the installed pwp sensors in order to track the SWRC in situ and to use a mixed dataset (wc+PH) in the parameters calibration because as seen in PART I the mixed dataset guarantees optimal results.

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my supervisor, prof Guido Gottardi, for the continuous support throughout the whole Ph.D study and research period, for his motivation, patience, enthusiasm and deep knowledge in the geotechnical engineering field. His guidance and advices have been invaluable to me.

Besides my advisor, I would like to thank my co-supervisor, dr Carmine Gerardo Gragnano, for his keen interest on me at every stage of the research. His suggestions, meticulous scrutiny, enthusiasm and dynamism have enabled me to complete this thesis.

I would like to extend my sincere gratitude to my second supervisor, prof. Olivier Buzzi who allowed me to work with him and his team in Newcastle University (Australia) providing me extensive personal and professional guidance in this part of my Ph.D research. I would like to express my sincere gratitude also to prof. Anna Giacomini for the kind help, generous support and comprehensive advices she gave me during this period abroad. I always remember these months with great affection and nostalgia.

My sincere thanks also go to prof Michele Calvello and prof. Riccardo Berardi for providing helpful feedback to improve this thesis and for opening discussions and comments to enrich my future research.

I am particularly grateful to prof. Michela Marchi with whom I have shared more than only research discussion, she showed me, by her example, what a good scientist (and person) should be.

I would like to acknowledge also the precious help of prof. Laura Govoni, especially in the first period of my Ph.D research in which she worked actively as supervisor of my research activity.

I would like to thank also dr Mario Marcolongo for the countless times in which he helped me fix the triaxial equipment and the laboratory everyday problems.

Special thanks go to prof. Irene Rocchi who followed me with enthusiasm and willingness in my first approach to geotechnical laboratory activities when I was a master student and she succeed in passing on me her passion.

My sincere thanks go to prof. Marco Bittelli for his precious expertise in soil and environmental physics which has been a valuable help expecially in the starting phase of the INFRASAFE project and in the first phase of the numerical modelling by means of Hydrus2D.

Nobody has been more important to me in the achievement of this important milestone than the members of my family. Their love and guidance are always with me in whatever goal I pursue and they are the ultimate role model of the person I want to be. Without such a team, I doubt that I would be in this place today. This thesis is dedicated to all of you.

Last but not least, I would like to thank all my special group of friends, Chiara, Bea, Ania, Giuly, Lisa, Marti, Leti, Anna. Thanks for always helping and encouraging me in pursuing my goals and being there in the most discouraging times. I have no valuable words to express how grateful I am to have all of you in my life.

BIBLIOGRAPHY (PART I; PART II)

Abbasi A., Annor F., van de Giesen N. (2015) *Developing a CFD-based approach to estimate evaporation from water surfaces in (semi-)arid regions*. EGU General Assembly,12-17 April, Austria.

Abbasi F. (2015). Advanced soil physics. Tehran University Publication, Iran.

Abbasi F., Simunek J., Feyen J., van Genuchten M.T., Shouse P.J.(2003). *Simultaneous inverse estimation of soil hydraulic and solute transport parameters from transient field experiments: homogeneous soil.* Transactions of the ASAE. 46(4): 1085–1095.

Abe K. and Ziemer R.R.(1991). *Effect of tree roots on a shear zone: modeling reinforced shear stress.* Can. J. For. Res. 21: 1012-1019.

Akaike, H. (1974). *A New Look at the Statistical Model Identification*. IEEE Transactions on Automatic Control, AC- 19, 716-723.

Ali F.H., Osman N., (2007). *Soil–roots composite: correlation between shear strength and some plant properties*. Electron. J. Geotech. Eng. V12 D.

Ali M.H., Abustan I. (2014). A new novel index for evaluating model performance. Journal of natural resources and development, 4.

Allen R.G., Pereira L..S., Smith M., Raes D., Wright J.L. (2005). *FAO-56 Dual Crop Coefficient Method for Estimating Evaporation from Soil and Application Extensions*. Journal of Irrigation and Drainage Engineering 131(1).

Allen R.G., Pereira L.S., Raes D., Smith M. (1998). *Crop evapotranspiration —guidelines for computing crop water requirements*. FAO Irrigation and drainage paper 56. Food and Agriculture Organization, Rome.

Allen R.G., Smith M., Pereira L.S., Perrier A. (1994). *An update for the calculation of reference evapotranspiration*. ICID Bulletin.

Arsenault R., Brissette F., Martel J-L. (2018). The hazard of split-sample validation in hydrological model calibration. Journal of Hydrology, 566.

Aslyng H.C. (1963). *Soil Physics terminology*. International Society of Soil Science Bulletin, 23: 7-10.

Babyak MA.(2004). *What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models*, Psychosomatic Medicine*,* 66:411-421.

Bacci L., Battista P., Rapi B., Sabatini F., Checcacci E.(2003). *irrigation control of container crops by means of tensiometers.* International Symposium on Managing Greenhouse Crops in Saline Environment 609: 467-474.

Belfiore G. and Urciuoli G. (2004). *Analisi del contributo meccanico delle radici alla resistenza del terreno* (*Interpretation of the root mechanical contribution to the soil shear strength*). In: Proceedings of the Annual Meeting of the Geotechnical Researchers 2004 – IARG 2004, Trento, Italy.

Bezerra Coelho C.R., Zhuang L., Barbosa M.C., Soto M.A.A., van Genuchten M.T. (2017). *Further tests of the HYPROP evaporation method for estimating the unsaturated soil hydraulic properties*. Journal of Hydrology and Hydromechanics, 66(2).

Bicalho K.V., Boussafir Y., Cui Y-J. (2018). *Performance of an instrumented embankment constructed with lime-treated silty clay during four years in Northest of France*. Transportation Geotechnics 17:100-116.

Bischetti G.B., Bonfanti F. and Greppi M. (2003). *Misura della resistenza alla trazione delle radici: apparato sperimentale e metodologia d'analisi*. Quaderni di Idronomia Montana 21/1.

Bishop A.W.(1955). *The use of the slip circle in the stability analysis of slopes*. Geotechnique 5(1):7-17.

Bittelli M. (2011). *Measuring soil water content: a review*. Hortechnology, 21(3).

Bittelli M., Campbell G.S., Tomei F.(2015). *Soil Physics with Python: transport in the Soil-Plant-Atmosphere System*, Oxford.

Bittelli M., Flury M.,(2009) *Errors in water retention curve determined with pressure plates*. Soil Science Society of America Journal, 73(5): 1453-1460.

Bittelli, M., Tomei, F., Pistocchi, A., Flury, M., Boll, J., Brooks, E. S., and Antolini, G. (2010). *Development and testing of a physically based, three-dimensional model of surface and subsurface hydrology*. Advances Water Resources, 33: 106–122.

Bitterlich S., Durner W., Iden S.C., Knabner P. (2004). *Inverse estimation of the Unsaturated Soil Hydraulic Properties from Column Outflow Experiments using free-form parameterization*. Vadose Zone Jounral, 3(3): 971-981.

Bordoni M., Bittelli M., Valentino R., Chersich S., Meisina C.(2017). *Improving the estimation of complete field soil water characteristic curves through field monitoring data*. Journal of Hydrology 552: 283-305.

Boso M., Romero E., and Tarantino A. (2003). *The use of different suction measurement techniques to determine water retention curves*. In Unsaturated soils: experimental studies. Edited by T. Schanz. Springer-Verlag, Berlin: 171–181.

Brooks R. H. and Corey A. T.(1964). *Hydraulic properties of porous media*, in Hydrology Paper, 3.

Bufon V.B., Lascano R.J., Bednarz C., Booker J.D., *Soil water content on drip irrigated cotton: Comparison of measured and simulated values obtained with the Hydrus 2D model,* Irrigation Science 30 (4).

Burger C.A., Shackelford C.D.(2001). *Soil-Water Characteristic Curves and Dual Porosity of Sand-Diatomaceous Earth Mixtures*. Journal of Geotechnical and Geoenvironmental Engineering, 127(9).

Burland J.B (1965) *Some aspects of the mechanical behaviour of partly saturates soils*. Proc. Conf. Moisture Equilibria and Moisture Changes in soil beneath Covered Areas: 270-278.

Burnham K.P. and Anderson D.R. (2004). *Multimodel Inference: Understanding AIC and BIC in Model Selection*. Sociological Methods & Research, 33(2), 261–304.

Cai G., Vanderborght J., Couvreur V., Mboh C.M. (2017). *Parametrization of RWU models considering dynamic root distributions and water uptake compensation*, Vadose Zone Journal 17(1).

Calabresi G., Colleselli F., Danese D., Giani G., Mancuso C., Montrasio L., Nocilla A., Pagano L., Reali E., Sciotti A. (2013). *Research study of the hydraulic behaviour of the Po River embankments*. Canadian Goetechnical Journal 50:947-960.

Caleffi C. & Cerutti, F. (2014). *Lavori urgenti per il miglioramento delle condizioni di stabilità del corpo arginale nei confronti dei fenomeni di filtrazione – fiume Secchia, località via Bozzala in Comune di Cavezzo (MO). Relazione geologica, geotecnica e sismica*. Report AIPO S2, 40 pp.

Campbell G.S., Smith D.M., Teare B.L. (2007). Application of a Dew Point Method to obtain the soil water characteristic. In Experimental Unsaturated Soil Mechanics, Editor Schanz T.,Springer, 71-77.

Campbell K.A., Hawkins C.D.B., (2003*). Paper birch and lodgepole pine root reinforcement in coarse-, medium-, and fine-textured soils*. Can. J. Forest Res. 33: 1580–1586.

Canadell J., Jackson R.B., Ehleringer J.R., Mooney H.A., Sala O.E., Schulze E.D. (1996), *Maximum rooting depth of vegetation types at the global scale,* Oecologia 108: 583-595.

Cardon G.E., Letey J. (1992) *Plant water uptake terms evaluated for soil water and solute movement models.* Soil science society of America Journal, 56(6).

Carrera J., Neuman S.P. (1986). Estimation of aquifer parameters under transient and steady state conditions:maximum likelihood method incorporating prior informations. Water Resources Research, 22(2): 199-210.

Caruso M., Jommi C. (2005). An evaluation of indirect methods for the estimation of hydraulic properties of unsaturated soils. Proceedings of International Conference on Problematic Soils, Eastern Mediterranean University, Famagusta, Cyprus, 25-27 May.

Casagrande A. (1937). *Seepage through dams*. New England Water Works, Vol. II, No. 2.

Ciollaro G., and N. Romano (1995). *Spatial variability of the soil hydraulic properties of a volcanic soil*, Geoderma, 65: 263-282.

Coelho E.F and Or D. (1999), Root distribution and water uptake patterns of corn under surface and subsurface drip irrigation, Plant and Soil, 206: 123-136.

Coelho E.F, Delfran B., dos Santos, Carlos A. V. de Azevedo (2007). *Sensor placement for soil water monitoring in lemon irrigated by micro sprinkler*. Bras. Eng. Agríc. Ambiental, 11(1): 46–52.

Coelho E.F., and Or. D. (1996) A parametric model for two-dimensional water uptake intensity by corn roots under drip irrigation, Soil Sci. Soc. Am. J., 60: 1039– 1049.

Cofie P., Koolen A.J. (2001). *Test speed and other factors affecting the measurement of tree root properties used in soil reinforcement models*. Soil Till. Res. 63: 51–56.

Coppin, N.J., Richards, I.G., 2007. *Use of vegetation in civil engineering*. CIRIA. LONDON.

Cunningham, M. (2000). *The mechanical behaviour of a reconstituted unsaturated soil*. Ph.D. thesis, Imperial College of Science, Technology and Medicine, London.

D Russo, E Bresler, U Shani, JC Parker (1991). *Analyses of infiltration events in relation to determining soil hydraulic properties by inverse problem methodology*, Water Resources Research, 27 (6): 1361-1373.

D Russo, G Bardhan, , D Goldstein, GJ Levy (2016), *Changes in the hydraulic properties of a clay soil under long-term irrigation with treated waste water*. Geoderma, 264: 1-9.

D'Alpaos L., Brath A., Fioravante V., Gottardi G., Mignosa P. and Orlandini S. (2014). *Relazione tecnico-scientifica sulle cause del collasso dell'argine del fiume Secchia avvenuto il giorno 19 Gennaio 2014 presso la frazione San Matteo*. Available online on http://www.regione.emilia-romagna.it.

Dapporto S., Rinaldi M., Casagli N., Vannocci P. (2003). *Mechanisms of riverbank failure along the Arno River, Central Italy*. Earth Surface Processes and Landforms 28(12): 1303–1323.

Day S.D. and Bassuk N.L. (1994). *A review of the effects of soil compaction and amelioration treatments on landscape trees*. Journal of Arboriculture 20(1): 9-17.

De Baets S., Poesen J., Reubens B., Wemans K.(2008). *Root tensile strength and root distribution of typical Mediterranean plant species and their contribution to soil shear strength*. Plant and Soil 305(1): 207-226.

Deb S.K., Shukla M.J., Simunek J., Mexal J.G. (2013) *Evaluation of Spatial and Temporal Root Water Uptake Patterns of a Flood-Irrigated Pecan Tree using the Hydrus (2D/3D) Model*, Journal of Irrigation and Drainage Engineering, 139(8).

Decagon Devices (2016a). *MPS-2 & MPS-6. Dielectric Water Potential Sensors*. Decagon Devices, Inc. Pullman, WA, US.

Decagon Devices (2016b). *GS3.Water Content, EC and Temperature Sensors*. Decagon Devices, Inc. Pullman, WA, US.

Delta-T Devices (2016). *SM150T. Soil Moisture Sensor*. Delta-T Devices, Cambridge, UK.

Dias A.S., Pirone M., Urciuoli G. (2017). *Review on the Methods for Evaluation of Root Reinforcement in Shallow Landslides*. Advancing Culture of Living with Landslides: 641-648.

Dirksen C. (1979). *Flux-Controlled Sorptivity Measurements to Determine Soil Hydraulic Property Functions*. SOIL SCI. SOC. AM. J., 43: 828-834

Doussan C., Pierret A., Garrigues E., Pagès L.(2006).*Water uptake by plant roots: II modelling of water transfer in the soil root-system with explicit account of flow within the root system – comparison with experiments*. Plant and Soil 283: 99–117.

Driessen P.M. (1986). The water balance of the soil. In modelling of agricultural production: weather, soils and crops, Editor Keulen H.V and Wolf J.,Pudoc,Wageningen, 76-116.

Driessen P.M., Konijn N.T.(1992). *Land-use systems analysis*. Wageningen: Wageningen Agricultural University.

Durner W.(1994). *Hydraulic conductivity estimation for soils with heterogeneous pore structure*. Water Resources Research, 30(2): 211-223.

Edward Washburn (1921). *The Dynamics of Capillary Flow*, 17.

Endo T., Tsuruta T. (1969). *The effect of the tree's roots on the shear strength of soil*. Annual report of the Hokkaido Branch Forest Experiment Station, Sapporo, Japan: 167-182.

Farthing M., Ogden F. (2017). *Numerical solution of Richards'equation: a review of advances and challenges*. Environmental Science, Soil Science Society of America Journal, 81(6).

Feddes R. A., Kowalik P. J., Malinka K. K., Zaradny H. (1978) *Simulation of Field Water Use and Crop Yield*. John Wiley and Sons, PUDOC Wageningen.

Feddes R.A., Bresler E., Neuman S.P. (1974). *Field test of a modified numerical model for water uptake by root systems.* Water resources research, 10(6).

Feddes R.A., Kowalik P., Kolinska-Malinka K., Zaradny H.,(1976) *Simulation of field water uptake by plants using a soil water dependent root extraction function.* Journal of Hydrology. 31(1-2): 12-26.

Ali F.H., N Osman (2007), *Soil-roots composite: Correlation between shear strength and some plant properties*. Electronic Journal of Geotechnical Engineering, 12.

Fredlund D.D., Rahardjo H., Fredlund M.D. (2012) *Unsaturated Soil Mechanics in Engineering Practice*. Wiley.

Fredlund D.G., Morgenstern N.R.,Widger R.A. (1978). *The shear strength of unsaturated soils*. Canadian Geotechnical Journal, 15(3):313-321.

Fredlund D.G., Xing A. (1994). *Equations for the soil-water characteristic curve*. Canadian Geotechnical Journal, 31(4).

Freeze R.A. (1971). *Three-dimensional, transient, saturated-unsaturated flow in groundwater basin*. Water Resources Research, 7(2): 347-366.

Frydman S., Operstein V. (2000), *The influence of vegetation on soil strength*, Proceedings of the Institution of Civil Engineers Ground Improvement, 4(2):81-89.

Gardner W.R. (1960). *Dynamic aspects of water availability to plants*, Soil science.

Gardner W.R. (1964). *Relation of root distribution to water uptake and availability*. Agron J 56: 41–45.

Gardner W.R., Miklich F.J. (1962). *Unsaturated conductivity and diffusivity measurements by a constant flux method*. Soil Science,93(4):271-274.

Garg A., Leung AK, CWW Ng (2015). *Comparison of soil suction induced by evapotranspiration and transpiration of S.heptaphylla*, Canadian Geotecnical journal, 52(12): 2149-2155.

Genet M., Stokes A., Salin F., Mickovski S.B., Fourcaud T., Dumail J.F, Van Beek R. (2005). *The influence of cellulose content on tensile strength in tree roots*. Plant Soil 278: 1-9.

Genet M.,Kokutse N., Stokes A., Fourcaud T., Cai X.Ji J., Mickovski S. (2008). *Root reinforcement in plantations of Cryptomeria japonica D. Don: effect of tree age and stand structure on slope stability*. Forest Ecology and Management.

Gerke H.H., van Genuchten M.Th. (1993a). *A dual-porosity model for simulating the preferential movement of water and solutes in structured porous media*. Water Resour. Res. 29 (2): 305–319.

Gerke H.H., van Genuchten M.Th., (1993b). *Evaluation of a first order water transfer term for variably saturated dual-porosity models*. Water Resour. Res. 29 (4): 1225–1238

Glinski J., Lipiec J. (1990) *Soil physical conditions and plant roots*, CRC-Press.

Gomes V.M.,Saraiva J.P, Lima B.S., Flores P.H.R, Assis A.O., Saraiva J.P., Gomes F.A., Wainer G.A., Magalhaes A.S., Calixto W., Flores P.R., Bulhoes J.S.(2017). *Analytical method for calculating the sensitivity index of system parameters*. CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)

Gottardi G. and Gragnano C.G. (2016). *On the role of partially saturated soil strength in the stability analysis of a river embankment under steady-state and transient seepage conditions*. 3rd European Conference on Unsaturated Soils E-UNSAT 2016, 9.

Gottardi G., Gragnano C.G., Rocchi I., Bittelli M. (2016). *Assessing River Embankment stability under transient seepage conditions*. VI Italian conference of researchers in Geotechnical Engineering - Geotechnical Engineering in Multidisciplinary Research: from Microscale to Regional Scale, CNRIG 2016.

Gragnano C.G., Bertolini I., Rocchi I., Gottardi G. (2019) *On the stability of a fully instrumented river embankment under transient conditions*. Geotechnical Research for Land protection and Development. Proceedings of CNRIG 2019.Springler.

Gragnano C.G., Gottardi G., Moscariello M., Cuomo S., Rocchi I. (2018). *Laboratory measurement of the mechanical and retention properties of a river embankment silty soil in partially saturated conditions*. 7th International Conference on Unsaturated Soils-Hong Kong University of Science and Technology, Hong Kong, 3-5 August.

Gray D.H., Ohashi, H., (1983). *Mechanics of fiber reinforcement in sand*. J. Geotech. Eng., ASCE 109 (3): 335–353.

Green S.R., Kirkham M.B., and Brent E.C. (2006). *Root uptake and transpiration: From measurements and models to sustainable irrigation*. Agric. Water Manage. 86: 165–176.

Greenway D.R. (1987). *Vegetation and slope stability*, Slope Stability, Wiley, Chichester:187-230.

Gupta H.V., Kling H., Yilmaz K., Martinez G.F. (2009). *Decomposition of the Mean Squared Error and NSE Performance Criteria: implications for improving hydrological modelling*. Journal of Hydrology 377(1): 80-91.

Hannan E.J. (1980). The estimation of the order of an ARMA process. *Ann. Statist*., 8(5):1071-1081

Hejazi, S.M., Sheikhzadeh, M., Abtahi, S.M. and Zadhoush, A. (2012). *A simple review of soil reinforcement by using natural and synthetic fibers*. Construction and Building Materials, 30:.100-116.

Hendriks M.R. (2010). *Introduction to Physical Hydrology*, Oxford University Press.

Hillel, D.(1980). *Fundamental of Soil Physics*. Academic Press, New York.

Hillel, D., Talpaz, H., van Keulen, H., 1976. *A macroscopic-scale model of water uptake by a nonuniform root system and of water and salt movement in the soil profile*. Soil Sci. 121, 242–255

Hirabayashi Y., Roobavannan M., Koirala S., Konoshima L. (2013). *Global flood risk under climate change*. Nature Climate Change 3(9): 816-821.

Hoogland J.C., Feddes R.A., Belmans C. (1981) *Root water uptake model depending on soil water pressure head and maximum extraction rate.* Acta Hortic 119:123-131

Hopmans J.W., Bristow K.L. (2002). *Current capabilities and future needs of root water and nutrient uptake modeling*. Adv. Agron. 77: 103–183.

Hopmans J.W., Simunek J., Romano N., Durner W.(2002). *Inverse Methods*. In Methods of Soil Analysis, 3.6.2: 963-1008.

Imbert C., Olchitizky E., Lassabatere T., Dangla P., Courtois A. (2005) *Evaluation of a thermal criterion for an engineered barrier system,* Engineering Geology, 81(3): 269-283

Ippisch O., Vogel H-J., Bastian P. (2006). *Validity limits for the van Genuchten-Mualem model and implications for parameter estimation and numerical simulation*. Advances in Water Resources 29(12) : 1780-1789.

Iverson R.M., Reid M.E. (1992).*Gravity-driven groundwater flow and slope failure potential: elastic effective stress model*. Water resource research, 28: 925-38.

Janbu N. (1954). *Stability analysis of slopes with dimensionless parameters*. Harvard soil mechanics series, 46(811).

Jarvis N. J. (1989) *A simple empirical model of root water uptake*. J. Hydrol., 107: 57–72.

Jarvis N.J.(2007) *A review of non-equilibrium water flow and solute transport in soil macropores: principles, controlling factors and consequences for water quality*. European journal of Soil Science 58(3): 523-546.

Javaux M., Couvreur V., Vander Borght J., Vereecken H. (2013). *Root water uptake: from three-dimensional biophysical processes to macroscopic modeling approaches*. Vadose Zone J., 12.

Javaux M., Schroeder T., Vanderborght J. and H. Vereecken. (2008). *Use of a three-dimensional detailed modelling approach for predicting root water uptake*. Vadose Zone Journal 7: 1079–1088.

Jiang H., Cai Y. Liu J. (2010). *Engineering properties of soils reinforced by short discrete polypropylene fiber*. Journal of Materials in civil Engineering, 22(12): 1315-1322.

Kling, H., Fuchs, M. and Paulin, M.(2012). *Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios* Journal of Hydrology: 424-425, 264–277.

Knight J. H. (1999). *Root distributions and water uptake patterns in Eucalyptus and other species.* Water and Salinity Issues in Agroforestry, Rural Ind. Res. and Dev.Corp, Rep. 99/37: 55–85.

Köhne J.M., Köhne S., Šimůnek J. (2009). A review of model applications for structured soils: water flow and tracer transport. J. Contam. Hydrol. 104: 4– 35.

Konikow L.F., Bredehoeft J.D. (1992). *Ground-water models cannot be validated*. Advances in water resources, 15(1): 75-83.

Kool J.B., Parker J.C. (1987). *Development and evaluation of closed-form expressions for hysteretic soil hydraulic properties*. Water Resources Research, 23, issue 1: 105-114.

Kosugi K.(1996). *Lognormal distribution model for unsaturated soil hydraulic properties*. Water Resources Research, 32(9):2697–2703.

Koumanov KS, Hopmans JW, Schwankl LW (2006). *Spatial and temporal distribution of root water uptake of an almond tree under microsprinkler irrigation*. Irrigation Science 24 (4).

Krause P., Boyle D., Base F. (2005). Comparison of different efficiency criteria for Hydrologic models. Advances in Geosciences 5: 89-97.

Lam L., Fredlund D.G. (1984). *Saturated-unsaturated transient finite element seepage model for geotechnical engineering*. Proceedings of the Fifth International Conference on Finite Elements in Water Resources, University of Vermont, Burlington, Vermont.

Lam L., Fredlund D.G. (1993). *A general limit equilibrium model for three-dimensional slope stability analysis.* Canadian Geotechnical Journal, 30(6).

Le Bourgeois O., Bouvier C., Brunet P., and Ayral P. A.(2016). *Inverse modeling of soil water content to estimate the hydraulic properties of a shallow soil and the associated weathered bedrock*. J. Hydrol. 541: 116–126.

Legates D., Mccabe G.J. (1999). *Evaluating the Use of "Goodness-of-fit" measures in hydrologic and hydroclimatic model validation*. Water Resources Research 35(1): 233-241.

Leib B.G., Caspari H.W., Redulla C.A., Preston K.A., Jalal J.J (2006). *Partial rootzone drying and deficit irrigation of Fuji apples in a semi-arid climate*. Irrigation Science 24: 85-99.

Leung A.K., Coo J.L., Ng C.W.W., Chen R. (2016). *New transient method for determining soil hydraulic conductivity function*. Canadian Geotechnical Journal. 53(8): 1332-1345.

Leung A.K., Garg A., Wai Ng C.W.(2015). *Effects of plant roots on soil-water retention and induced suction in vegetated soil*, Engineering Geology, 193: 183-197.

Li S.L.; Liang W.L.(2019) *Spatial–Temporal Soil Water Dynamics beneath a Tree Monitored by Tensiometer-Time Domain Reflectometry Probes*. Water, 11, 1662.

Liang Y., Chen J., Chen L. (2011) *Mathematical Model for Piping erosion based on fluid-solid interaction and soils structure*. GeoHunan International Conference 2011, June 9-11, Hunan, China.

Likos W., Lu N., Godt J.W. (2014). *Hysteresis and Uncertainty in Soil Water-Retention Curve Parameters.* Journal of Geotechnical and Geoenvironmental Engineering 140(4).

Liu H., Lennartz B. (2018). *Hydraulic properties of peat soils along a bulk density gradient—A meta study*. Hydrological Processes, 33: 101-114.

Liu J., Chen P., Li W., (2018). *Assessing hydraulic hysteresis models to characterize unsaturated flow behaviour under drying and wetting conditions*. International Journal of Geomechanics. 18(7)

Liu K., Vardon P.J., Arnold P., Hicks M.A. (2015). *Effect of hysteresis on the stability of an embankment under transient seepage*. International Symposium on Geohazards and Geomechanics (ISGG2015), IOP Conf. Series: Earth and Environmental Science, 26.

Liu K., Vardon P.J., Hicks M.A., Arnold P.(2016) *Combined effect of hysteresis and heterogeneity on the stability of an embankment under transient seepage*. Engineering Geology.

Liu K.F., Yang X.R., Xie X.Y., Wu C.F., Liu Y.H. (2011*). Laboratory triaxial test study on soil reinforced with roots of Manilagrass*. Advanced materials research, 250-253.

Liu X., Buzzi O. (2014). *Use of Hand-Spray Plaster as a coating for soil bulk volume measurement*. Geotechnical Testing Journal, 37(3).

Lloret A., Villar M.V.(2007). *Advances on the knowledge of the thermo-hydro-mechanical behaviour of heavily compacted "FEBEX" bentonite*. Physics and Chemistry of the Earth, 32(8–14): 701-715

Loague K. and Green R.E. (1991). Statistical and graphical methods for evaluating solute transport models: overview and application*. Journal of Contaminant Hydrology*, 7: 51-73.

Lourenco S.D.N., Gallipoli D., Toll D.G., Evans F., Medero G.M. (2007). Determination of the Soil Water Retention Curve with Tensiometers. In Experimental Unsaturated Soil Mechanics.

Lu N., Kaya M. (2013). A drying cake method for measuring suction stress characteristic curve, soil water retention and suction-stress characteristic curves. Journal of Geotechnical and Geoenvironmental Engineering, 140(5): 1-10.

Lu N., Likos W.J. (2004). *Unsaturated soil mechanics*. Wiley.

Lu N., Likos W.J. (2006). *Suction stress characteristic curve for unsaturated soil*. Journal of Geotechnical and Geoenvironmental engineering 132(2): 131-142.

Lu N., Sener-Kaya B., Wayllace A., Godt J.W. (2012). *Analysis of rainfall-induced slope stability using a field of local factor of safety*. Water resource research, 48.

Lu N., Wayllace A., Chen P. (2019). *The Slope Cube Module for Hydrus 2D/3D: simulating Slope Stress and Stability in Variably-Saturated Hillslopes*. Manual. Version 1.0

Schwarz M, Cohen D, Or D (2011) *Pullout tests of root analogs and natural root bundles in soil: Experiments and modeling*, Journal of Geophysical Research: Earth Surface, 116

Maffra C, Sousa R, Sutili F and Pinheiro R (2019). *The Effect of Roots on the Shear Strength of Texturally Distinct Soils*. Floresta e Ambiente, 26(3): 3-5.

Mashayekhi P., Dashtaki S.G., Mosaddeghi M.R., Shirani H., Nodoushan A.R.M.(2016). *Different scenarios for inverse estimation of soil hydraulic parameters from double-ring infiltrometer data using HYDRUS 2D/3D.* International agrophysics 30(2): 203-210.

Mendes J., Toll D.G., Augarde C.E.(2008). *A system for field measurement of suction using high capacity tensiometers*. Proceedings of the first European conference on unsaturated soils E-UNSAT2008, Durham, UK. Unsaturated Soils: Advances in Geo-Engineering, Taylor and Francis Group.

Mickovski S.B., F.Bransby, P.H.Hallett (2009) *Mechanical Reinforcement of Soil by Willow Roots: Impacts of root properties and root failure mechanism*, Soil Science Society of America Journal, 73(4).

Mickovski S.B., Van Beek L.P.H., (2009). *Root morphology and effects on soil reinforcement and slope stability of young vetiver (Vetiveria zizanioides) plants grown in semi-arid climate*. Plant Soil, 324:43–56.

Miller, C.J., Mi, H., Yesiller, N.(1998). *Experimental Analysis of Desiccation Crack Propagation in Clay Liners*. Journal of the American Water Resources Association 34 (3): 677-686.

Miranda Neto M.I., Mahler C.F.(2017*) Study of the Shear Strength of a Tropical Soil with Grass Roots*. Soils and Rocks, 40(1):31-37.

Molz F.J. (1981). *Models of water transport in the soil-plant system: A review*. Water Resources Research, 17(5).

Monteith J.L. (1965) *Evaporation and environment*. Proceedings Symposium of the Society for Experimental Biology, 19:205–234.

Morgan R.P.C, Rickson R.J (2003). *Slope stabilization and Erosion Control: a bioengineering approach*. Taylor and Francis.

Morgenstern N.R. and Price V.E. (1967). *A numerical method for solving the equations of stability of general slip surfaces*. Computer Journal, 9: 388-393.

Mualem Y. 1976. *A new model predicting the hydraulic conductivity of unsaturated porous media*. Water Resources Research 12: 513–522.

Murphy A.H.(1988) *Skill scores based on the mean square error and their relationships to the correlation coefficient*. Monthly weather review, 116(12).

Naasz R., Michel J.C., Charpentier S.(2005). *Modeling oxygen and water flows in peat substrate with root uptakes*. Proceedings of the International Symposium on Growing Media, Angers (FR), France. International Society Horticultural Science, 779:191-197.

Nash J.E. and Sutcliffe J.V. (1970). *River Flow Forecasting through Conceptual Model*. Part 1—A Discussion of Principles. Journal of Hydrology, 10: 282-290.

Ni J., Wai Ng C.W., Gao Y. (2020). *Modelling root growth and soil suction due to plant competition*, Journal of Theoretical Biology, 484.

Nielsen D. R., Luckner L. (1992). *Theoretical aspects to estimate reasonable initial parameters and range limits in indentification procedures for soil hydraulic properties*. In, Proc. Intl. Workshop on Indirect

Methods for Estimating the Hydraulic Properties of Unsaturated Soils, edited by M. Th van Genuchten, F. J. Leij, and L. J. Lund, University of California, Riverside: 147-160.

Nodichao L., Chopart J.L, Roupsard O., Vauclin M. (2011) *Genotypic variability of oil palm root system distribution in the field. Consequences for water uptake*. Plant and Soil 341(1): 505-520.

O'Loughlin C.L. (1981). *Tree roots and slope stability. What's new in Forest Research*. Forest Research Institute Publication No. 104, Rotorua, New Zealand.

Osman N., Abdullah M.N, Abdullah C.H. (2011). *Pull-out and tensile strength properties of two selected tropical trees.* Sains Malaysiana 40(6): 577-585.

Peduzzi P., Dao H., Herold C., Mouton F (2009). *Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index*. Natural Hazards and Earth System Sciences, 9: 1149-1159.

Penman H.L. (1948) *Natural evaporation from open water, bare soil and grass*. Proceedings of the Royal Society of London, 193:120–145.

Peterson, D.L., D.M. Glenn, and S.D. Wolford (1992). *Tensiometer-irrigation control valve*. Appl. Eng. Agric. 9: 293–297.

Pham H.Q., Fredlund D.G., Barbour S.L. (2005). *A study of hysteresis models for soil-water characteristic curves*. Canadian Geotechnical Journal, 42 (6).

Phogat V., Skewes M.A., Cox J.W., Simunek J. (2016). *Statistical assessment of a numerical model simulating agro hydro-chemical processes in soil under drip irrigated mandarin tree*. Irrigation & Drainage Systems Engineering, 5(1).

Pollen N., Simon A. (2005). *Estimating the mechanical effects of riparian vegetation on stream bank stability using a fiber bundle model*. Water Resources Research, 41.

Raats P.A.C. (1974). *Steady flows of water and salt in uniform soil profiles with plant roots*. Soil Sci. Soc. Am. Proc. 38: 717–722.

Radcliffe D.E., Simunek J.J.(2010). *Soil physics with Hydrus: Modeling and applications*.CRC-Press.

Rahardjo H., Kim Y., Satyanaga A.(2019) *Role of unsaturated soil mechanics in geotechnical engineering*. International Journal of Geo-Engineering 10(8).

Raue A., Kreutz C., Maiwald T., Bachmann J., Schilling M., Klingmuller U., Timmer J. (2009). *Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood.* Bioinformatics 25(15): 1923-1929.

Richards L.A.(1931) *Capillary Conduction of Liquids through Porous Medium*. Journal of Applied Physics,1 : 318-333.

Rinaldi M., Casagli N.(1999). *Stability of streambanks formed in partially saturated soils and effects of negative pore water pressure: the Sieve River (Italy)*. Geomorphology, 26, issue4: 253-277.

Rinaldi M., Casagli N., Dapporto S., Gargini A. (2004). *Monitoring and modelling of pore water pressure changes and riverbank stability during flow events*. Earth Surface Processes and Landforms 29: 237-254.

Robbins N.E, Dinneny J.R (2015). *The divining root: moisture-driven responses of roots at the micro and macro scale*. Journal of experimental botany, 66(8): 2145-2154.

Robertson, P.K. (2009). *Interpretation of cone penetration tests - a unified approach*. Canadian Geotechnical Journal, 46(11): 1337-1355.

Rocchi I., Gragnano C.G., Gottardi G., Govoni L., Bittelli M. (2018a). *In situ measurements of soil water content and suction to assess river embankment stability under transient flow conditions*.7th International Conference on Unsaturated Soils-Hong Kong University of Science and Technology, Hong Kong, 3-5 August

Rocchi I., Gragnano C.G., Govoni L., Mentani A., Bittelli M., Castiglione P., Buzzi O., Gottardi G. (2018b). *A new technique for deep in situ measurements of soil water retention behaviour*. Geotechnical Research, vol 5, issue GR1.

Romano N., Santini A. (1999). *Determining Soil Hydraulic Functions from Evaporation Experiments by a parameter estimation approach: experimental verifications and numerical studies*. Water Resources Research 35(11):3343-3359.

Romero E.,Gens A., Lloret A.(1999). *Water permeability, water retention and microstructure of unsaturated compacted Boom clay*. Engineering Geology, 54(1–2): 117-127.

Saraiva J.P., Lima B.S., Gomes V.M., Flores P.H.R., Gomes F.A., Assis A., da Cunha Reis M.R., Araujo W.R.H., Abrenhosa C., Calixto W.P. (2017*) Calculation of sensitivity index using one-at-a-time measures based on graphical analysis*. 18th International Scientific Conference on Electric Power Engineering (EPE).

Scanlan C.A., Hinz C.(2010). *Insight into the processes and effects of rootinduced changes to soil hydraulic properties*. 19th World congress of soil science, soil solutions for a changing world, 1-6 August, Brisbane, Australia.

Schaap M.G., Leij F.J., van Genutchen M.T. (2001). *Rosetta: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions*. Journal of Hydrology, 251(3-4): 163-176.

Schelle H., Iden S.C., Peters A., Durner W. (2010). *Analysis of the agreement of soil hydraulic properties obtained from multistep-outflow and evaporation methods*. Vadose Zone Journal 9(4): 1080-1091.

Schiechtl H.M, Stern R. (1996) *Ground bioengineering techniques for slope protection and erosion control,* John Wiley and Sons.

Schwarz M., Cohen D., Or D. (2010). *Root-soil mechanical interactions during pullout and failure of root bundles.* Journal of Geophysical Research: Earth Surface, 115 (F4).

Schwarz M., Giadrossich F., Cohen D.(2013) *Modeling root reinforcement using a root-failure Weibull survival function.* Hydrology and Earth System Sciences, 17: 4367-4377.

Schwarz M., Rist A., Cohen D., Giadrossich F., Egorov P., Buttner D., Stolz M., Thormann J.J. (2015). *Root reinforcement of soils under compression*. Journal of Geophysical Research: earth surface. 120 (10).

Scott P.S., Farquhar G.J., Kouwen N. (1983). *Hysteretic effects on net infiltration*, In Advances in Infiltration, American Society of Agricultural Engineers Publication:11-83.

Shah N., Ross M., Ladde G.S. (2008). *Dynamic modelling of root water uptake using soil moisture data*. Neural, Parallel and Scientific Computations 16(1): 105-124.

Sharp R.E., Davies W.J. (1985) *Root Growth and Water Uptake by maize plants in drying soil*. Journal of Experimental Botany, vol 36, issue 9: 1441-1456.

Sidle R.C., Bogaard T. (2016). *Dynamic earth system and ecological controls of rainfall-initiated landslides*, Eart-Science Reviews 159.

Simon A., Curini A., Darby S.E., Langendoen E.J. (2000). *Bank and near-bank processes in an incised channel*. Geomorphology 35:197-217.

Simunek J., de Vos J.A.(1999). *Inverse optimization, calibration and validation of simulation models at the field scale*. Modelling of transport processes in soils at various scales in time and space; international workshop of EurAgEng's field of interest on soil and water. Wageningen, Wageningen Pers. Editors J.Feyen, K.Wiyo, 1999: 431-445.

Simunek J., Hopmans J.W. (2002). 1.7 *Parameter Optimization and Nonlinear fitting*. Methods of Soil Analysis, part 4 Physical Methods, editors J.H.Dane, G.C.Topp.

Šimůnek J., Jarvis N. J., van Genuchten M. Th., Gärdenäs A. (2003). *Review and comparison of models for describing nonequilibrium and preferential flow and transport in the vadose zone*. J. Hydrol. 272 (1-4): 14-35.

Simunek J., Sejna M., van Genuchten M.T. (2006). *The HYDRUS Software Package for Simulating Two- and Three-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media*. User Manual, Version 1.0. University of California-Riverside Research Reports. 161 pp.

Šimůnek J., van Genuchten M. Th., Šejna M. (2005). *The HYDRUS-1D software package for simulating the onedimensional movement of water, heat, and multiple solutes in variably saturated media*. Version 3.0. HYDRUS Software Series 1. Riverside, Cal.: University of California, Department of Environmental Sciences.

Simunek J., van Genuchten M.T., Sejna M. (2012). *Hydrus: model use, calibration and validation*. Transactions of the ASABE. 55(4): 1261-1274.

Simunek J., van Genuchten M.T., Wendroth O. (1998). *Parameter estimation analysis of the evaporation method for determining soil hydraulic properties*. Soil science society of America Journal, 62(4):894-905.

Simunek J., van Genuchten M.Th., Sejna M. (2012a). Hydrus: model use, calibration and validation. Transactions of the ASAE. American Society of Agricultural Engineers 55: 1261-1274.

Simunek J., van Genuchten M.Th., Sejna M. (2012b). *The HYDRUS Software Package for Simulating the Two- and Three-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Porous Media*. Technical manual, version 2.

Simunek J., Wendroth O., van Genuchten M.T. (1999). *Estimating unsaturated soil hydraulic properties from laboratory tension disc infiltrometer experiments*. Water Resources Research. 35(10):2965-2979.

Simunek J., Wendroth O., Wyper N., van Genuchten M.T. (2001). *Non-equilibrium water flow characterized from an upward infiltration experiment*. European Journal of Soil Science 52: 13-24.

Simunek J.J., Hopmans J.W. (2009). *Modeling compensated root water and nutrient uptake*. Ecological Modelling 220(4): 505-521.

Šimůnek, J. and van Genuchten M. Th. (2008) *Modeling nonequilibrium flow and transport with HYDRUS*, Vadose Zone Journal, Special Issue "Vadose Zone Modeling", 7(2): 782-797.

Simunek, J., H. Othmer, and M. Th. van Genuchten, (1996) *HYDRUS-2D, Simulation Water Flow and Solute Transport in Two-Dimensional Variably Saturated Media*, TPS 53, International Groundwater Modeling Center, Colo. School of Mines, Golden,CO.

Simunek, J., Van Genuchten, M. T., and Sejna, M. (2009) *The HYDRUS1-D software package for simulating the one-dimensional movement of water, heat, and multiple solutes in variably-saturated media*, University of California-Riverside Research Reports, 3:1–240.

Šimůnek, J., Wendroth O., van Genuchten M. Th. (1998). *A parameter estimation analysis of the evaporation method for determining soil hydraulic properties*. SSSA J. 62(4): 894-905.

Singh J., Knapp H.V., Demissie M.(2004). *Hydrologic modelling of the Iroquois river watershed using HSPF and SWAT*. Journal of the American Water Resources Association (JAWRA),41(2): 343-360.

Sivapatham P., Kumar A.A., Fares A. (2000). An evaluation of soil water status using tensiometers in a sandy soil profile under citrus production. Soil Science 165(4): 343-353.

Somma F., Hopmans J.W., Clausnitzer V. (1998) *Transient Three-dimensional modelling of soil water and solute transport with simultaneous root growth, root water and nutrient uptake*, Plant and Soil, 202: 281-293.

Spencer E., Tech M.S. (1967). *A method of analysis of the stability of embankments assuming parallel inter-slice forces*. Geotechnique,17: 11-26.

Stokes A., Ball J., Fitter A.H., Brain P., Coutts M.P. (1996). *An experimental investigation of the resistance of Model Root System to Uprooting*. Annals of Botany, 78 (4): 415-421.

Stokes A., Douglas G.B., Fourcaud T., Giadrossich F., Gillies C., Hubble T., Kim J.H., Loades K.W., Mao Z., Mclvor I.R., Mickovski S.B., Mitchell S., Osman N., Phillips C., Poesen J., Polster D., Preti F., Raymond P., Rey F., Schwarz M., Walket L.R. (2014). *Ecological mitigation of hillslope instability: ten key issues facing researchers and practitioners*, Plant and Soil, 377: 1-23.

Stone M. (1979). *Comments on model selection criteria of Akaike and Schwarz*. Journal of the Royal Statistical Society, 41: 276–278.

Sun D. A., Sun W., and Xiang, L. (2010). *Effect of degree of saturation on mechanical behaviour of unsaturated soils and its elastoplastic simulation*. Computers and Geotechnics, 37(5): 678–88.

Tanoue M., Hirabayashi Y., Ikeuchi H. (2016). *Global-scale flood vulnerability in the last 50 years*. Scientific Reports, 6.

Tarantino A.(2013) *Basic concepts in the Mechanics and Hydraulics of Unsaturated Geomaterials*. In Mechanics of Unsaturated Geomaterials by Lyesse Laloui, Wiley.

Taylor R.L. and Brown C.B. (1967). *Darcy Flow with a Free Surface*. Journal of the Hydraulic Division, 93(2): 25-33.

Taylor S.A.; Ashcroft G.L.(1972). *Physical Edaphology: The Physics of Irrigated and Nonirrigated Soils*; W.H. Freeman: San Francisco, CA, USA, 532.

Thirel G., Andreassian V., Perrin C. (2015). *On the need to test hydrological models under changing conditions*. Hydrological Sciences Journal 60(7).

Thomas R. E., Pollen-Bankhead N., (2010). *Modeling root-reinforcement with a fiber-bundle model and Monte Carlo simulation*. Ecological Engineering, 36: 47–61.

Tokunaga, T.K. (2009). *Hydraulic properties of adsorbed water films in unsaturated porous media*. Water Resources Research, 45(6).

Toll D.G., Lourenco S.D.N, Mendes J., Gallipoli D., Evans F.D., Augarde C.E., Cui Y.J., Tang A.M., Rojas J.C., Pagano L., Mancuso C., Zingariello C., Tarantino A. (2016). *Soil suction monitoring for landslides and slopes*. Quarterly Journal of Engineering Geology and Hydrogeology 44: 23-33.

Tosi M. (2007). Root tensile strength relationships and their slope stability implications of three shrub species in Northern Appennines (Italy). Geomorphology 87(4): 268-283.

Trigila A., Iadanza C., Bussettini M., Lastoria B. (2018). *Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio- Rapporto 2018*. ISPRA.

Trigila A., Iadanza C., Bussettini M., Lastoria B., Barbano A.(2015). *Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio- Rapporto 2015*. ISPRA.

UMS (2011). *T8. Long-term Monitoring Tensiometer*. UMS GmbH, München Germany.

Vaciago G. (2013). *The SafeLand compendium of landslide risk mitigation measure, in* MARGOTTINI C., CANUTI P., SASSA K. (eds.), *Landslide Science and Practice, 6,* Springer-Verlag, Berlin-Heidelberg: 683-689.

Vacondio R., Aureli F., Ferrari A., Mignosa P., Dal Palù A. (2015). *Simulation of the January 2014 flood on the Secchia River using a fast and high-resolution 2D parallel shallow-water numerical scheme*. Natural Hazard 80:103-125.

van Genuchten M.(1980). *A closed-form equation for predicting the Hydraulic conductivity of unsaturated soils*. Soil Science Society of America Journal 44(5).

van Genuchten M.Th.,Leij F.J.,Yates S.R. (1991).*The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils*, Version 1.0. EPA Report 600/2-91/065, U.S. Salinity Laboratory, USDA, ARS, Riverside, California.

van Genuchten. M.T (1987). *A numerical model for water and solute movement in and below the root zone*. Research Rep. 121, U.S. Salinity Laboratory, USDA, ARS, Riverside, CA.

Verbist K.J. Baetens W.M. Cornelis D., Gabriels C.,Torres and G. Soto (2009). *Hydraulic conductivity as influenced by stoniness in degraded drylands of Chile*, Soil Science Society of America Journal, 73: 471– 484.

Veylon G, Ghestem M, Stokes A, Bernard A (2015). Quantification of mechanical and hydric components of soil reinforcement by plant roots. Can Geotech J, 52: 1839–1849.

Vogel T., Cislerova M. (1988). *On the reliability of unsaturated hydraulic conductivity calculated from the moisture retention curve*. Transport in Porous Media 3(1):1-15.

Vogel T., van Genuchten M., Cislerova M. (2000). *Effect of the shape of soil hydraulic functions near saturation on variably-saturated flow predictions.* Advances in Water Resources 24(2):133-144.

Vrugt J., Wijk M.V.,.Hopmans J.W, Simunek J.J. (2001). *One, Two, Three dimensional Root Water Uptake Functions for Transient Modeling,* Water Resources Research, 37(10): 2457-2470.

Vrugt J.A., Bouten W., Weerts A.H. (2001) *Information content of data for identifying soil hydraulic parameters from outflow experiments*. Soil science society of America Journal 65(1):19-27.

Vrugt J.A., Hopmans J.W., Šimůnek J. (2001). C*alibration of a two-dimensional root water uptake model.* Soil Sci. Soc. Am. J. 65(4): 1027-1037.

Waldron L.J. (1977). *The shear resistance of root-permeated homogeneous and stratified soil*. Soil science society of America Journal, 41(5).

Waldron L.J., Dakessian S.(1982). *Effect of grass, legume and tree roots on soil shearing resistance*. Soil Sci. Soc. Am. J., 46: 894-899.

Waldron L.J., Dakessian S., Nemson J.A. (1983). *Shear Resistance Enhancement of 1.22 –Meter Diameter Soil Cross sections by Pine and Alfalfa roots*. Soil science society of America Journal. 47(1).

Ward P.J., Jongman B., Weiland F.S., Bouwman A., Beek R.V., Bierkens M.F.P, Ligtvoet W., Winsemius H.C (2013). *Assessing flood risk at the global scale: model setup, results and sensitivity*. Environmental Research Letters, 8.

Watabe Y., Leroueil S., and Le Bihan I.P. (2000). *Influence of compaction conditions on pore-size distribution and saturated hydraulic conductivity of a glacial till*. Canadian Geotechnical Journal., 27: 761–773.

Watson, K.K. (1966). *An instantaneous profile method for determining the hydraulic conductivity of unsaturated porous materials*. Water Resources Research, 2(4), 709-715.

Wesseling P. (1991) *An Introduction to Multigrid Methods*. John Wiley and Sons, New York.

Whisler F. D., Klute A., Millington R.J. (1968). *Analysis of Steady-State Evapotranspiration from a Soil Column*, Soil Science Society of America Journal, 32(2).

Willmott C.J. (1982). *Some comments on the evaluation of model performance*. Bulletin American Meteorological Society, 63: 1309-1313.

Willmott C.J., Ackleson S.G., Davis R.E., Feddema J.J., Klink K.M., Legates D.R., O'Donnell J., Rowe C.M. (1985). *Statistics for the evaluation and comparison of models*. Journal of Geophysical Research. 90(C5): 8995-9005.

Willmott CJ. (1981). *On the validation of models*. Physical Geography 2: 184– 194.

Wind G.P. (1968). *Capillary conductivity data estimated by a simple method*. In: Rijtema, P.E., Wassink, H. (Eds.), Water in the Unsaturated Zone, vol. 1. Proceedings of the Wageningen Symposium, 19–23 June 1966. Int. Assoc. Sci. Hydrol. Publ. (IASH), Gentbrugge, The Netherlands and UNESCO, Paris.

Winsemius H., Aerts J.C.H., van Beek L.P.H., Bierkens M.F.P. (2015). *Global drivers of future river flood risk*. Nature Climate Change, 6(4).

Wu J., Liu Z., Chen D., Huang G., Zhou L., Fu S.(2011). *Understory plants can make substantial contribution to soil respiration: evidence from two subtropical plantations*. Soil Biology and Biochemistry, 43 (11): 2355-2357.

Wu J., Zhang R., Gui S. (1999). *Modeling soil water movement with water uptake by roots*. Plant and Soil, 215 (7-17).

Wu T.H. (1976). *Investigation on landslides on Prince of Wales Island*. Alaska Geotech. Rpt. No 5, Dpt. Of Civil Eng., Ohio State Univ., Columbus, USA.

Wu T.H., Beal P.E., Lan C. (1988). *In-situ shear test of soil–root systems*. J. Geotech. Eng. (ASCE) 114: 1376–1394.

Wu T.H., McKinnell W.P., Swanston D.N. (1979). *Strength of tree roots and landslides on Prince of Wales Island, Alaska*. Canadian Geotechnical Journal 114(12): 19-33.

Wu T.H., Watson A. (1998). *In situ shear tests of soil blocks with roots*. Can. Geotech. J. 35 (4): 579–590.

Wu W., Switala B., Acharya M.S., Tamagnini R. (2015). *Effect of vegetation on stability of soil slopes: numerical aspect*, Springer Series in Geomechanics and Geoengineering: 163-177.

Wythers K.R., Lauenroth W.K., Paruelon J.M. (1999). *Bare-Soil Evaporation Under Semiarid Field Conditions*. Division Soil & Water Management & Conservation.

Mualem Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resour. Res. 12: 513–522.

Yang Y., Chen L. (2016). *Effect of root moisture content and diameter on root tensile properties*. PLoS ONE 11(3).

Zhang C., Chen L., Jiang J. (2014). *Why fine tree roots are stronger than thicker roots: The role of cellulose and lignin in relation to slope stability*. Geomorphology, *206*: 196-202.

Zhang C., Chen L.H., Liu Y.P., Liu X. (2010). *Triaxial compression test of soil-root composites to evaluate influence of roots on soil shear strength*. Ecological Engineering 36(1): 19-26.

Zhu H., Zhang L.M. (2015). Evaluating suction profile in a vegetated slope considering uncertainty in transpiration. Comput. Geotech., 63: 112-120.

Zhu H., Zhang L.M., Xiao T., Li X. (2017). *Enhancement of slope stability by vegetation considering uncertainties in root distribution*. Computers and Geotechnics 85: 84-89.

Zhuang L., Coelho C.R.B., Hassanizadeh S.M., van Genuchten (2017). *Analysis of the Hysteretic Hydraulic Properties of Unsaturated Soil*. Vadose Zone Journal, 16(5).

Ziemer R. (1981). *The role of vegetation in the stability of forested slopes*. Proceedings of the Christchurch Symposium Volume: Int. Assn. Hydrol. Sci. Pub. No. 132: 343-361.

Zuo Q., Jie F., Zhang R., Meng L. (2004). *A generalized function of wheat's root length density distributions*, Vadose Zone J.,3: 271–277.

Zuo Q., Shi J, Li Y., Zhang R (2006). *Root length density and water uptake distributions of winter wheat under sub-irrigation*. Plant Soil 285: 45–55.